THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIV March, 1933 Number 3

SEX DIFFERENCES IN MENTAL GROWTH AND DECLINE[1]

H. S. CONRAD, H. E. JONES AND H. H. HSIAO

Institute of Child Welfare, University of California

The present report is based on data from the Army Alpha intelligence examination, administered to five hundred eighty-one males and six hundred seven females in a representative rural group, between the ages of ten and sixty. Details of sampling and of test administration are given in a previous report (7).

Prior work has shown that for the total sample, the developmental curve is linear to about sixteen years, with a negative acceleration beyond sixteen to a peak between the ages of eighteen and twenty-one. A decline follows, which is much more gradual than the rise during adolescence, but which by the age of fifty-five involves a recession to the fourteen-year level (as computed from a fitted or a smoothed curve (3, 7)).

We are concerned here with the comparative analysis of the developmental data for the two sexes. Table I gives the number of cases in each age-sex group; Tables II and III present mean and median scores, respectively, for these groups. Means and medians are presented not only for performance in the total Alpha, but also for the "abbreviated Alpha,"[2] and for each subtest of the Alpha.[3] Table

[1] The data for the present study were collected with the aid of a research grant from the Social Science Research Council of Columbia University. Acknowledgments are due to Professors R. S. Woodworth and A. T. Poffenberger, of Columbia University, for advice and administrative assistance.

[2] The "abbreviated Alpha" is the same as total Alpha, except for the elimination of the scores in subtest 8 (which measures "general information") and subtest 4 (which measures vocabulary through an opposites test). Due to the fact that at the upper ages, accumulated experience tends unduly to weight the information tests 4 and 8, we have considered the abbreviated Alpha to provide a better com-

IV gives the reliability of sex differences in each of these tests, for combined age-groups. From Table IV it can be seen that while the males are fairly consistently inferior in total score, the difference is reliable only in the age period ten to fifteen. In later adolescence the boys tend to catch up in mental test ability (in this respect the mental growth curves are similar to physical growth curves). Minor selection differences may influence the results at certain ages; for example, it appears likely that a slightly inferior selection of women was obtained at ages twenty-five to twenty-nine, due to the policy, in home canvassing, of including all families with two or more children (7).[1] In Tables II–IV, a slight but unreliable tendency appears for sex differences to diminish to age forty or forty-five, and to increase thereafter. This tendency, consistent in nearly all of the subtests, could be determined by errors of sampling.[2]

TABLE I.—NUMBER OF CASES IN EACH AGE-GROUP

Sex	Age-group									
	10	11	12	13	14	15	16	17	18	
Male	16	22	34	35	27	31	42	29	27	
Female	18	33	31	35	29	30	32	21	19	
	19–21	22–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	Total
Male	46	23	41	51	36	46	27	32	16	581
Female	41	21	47	55	59	51	41	27	17	607

parative measure of the intelligence of different age-groups, than the total Alpha test (2, 7).

[3] There are eight Alpha subtests, which may be briefly described as follows: (1) Oral directions; (2) arithmetical problems; (3) "common sense"; (4) opposites or vocabulary; (5) disarranged sentences; (6) numerical completions; (7) verbal analogies; (8) general information. For more detailed description, see reference No. 11.

[1] We have shown elsewhere (5) that in this rural sample, lower intelligence is associated with earlier childbirth, although not with the total size of completed families.

[2] We should not, however, overlook the possibility that the growing discrepancy between males and females in the upper ages may be real. Using correlations between age and score, both Carter (1, p. 44) and Price (9, pp. 12, 58–59) have observed a slightly more rapid decline among men than among women.

TABLE II.—MEAN RAW SCORES OF FIVE HUNDRED EIGHTY-ONE MALES (M), AND SIX HUNDRED SEVEN FEMALES (F)

Age-group	Total alpha		Abbreviated alpha*		Alpha subtest number													
					1		2		3		4		5		6		7		8	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
10	41.3	47.2	31.3	34.7	3.8	3.6	4.9	5.1	4.5	3.9	3.8	6.3	6.4	6.1	4.5	5.3	8.8	9.7	8.3	8.1
11	43.6	48.9	30.9	35.2	4.8	4.6	5.6	5.3	3.9	3.9	4.9	5.7	6.5	6.6	5.0	5.4	7.2	9.4	8.1	8.6
12	54.1	59.5	38.4	42.2	4.9	5.5	6.7	6.1	4.8	5.6	5.8	7.1	7.7	9.3	6.4	5.3	10.3	12.0	10.4	11.7
13	54.4	67.9	38.2	49.8	4.9	6.0	6.9	6.8	5.8	6.1	6.9	8.0	8.1	10.6	5.2	7.4	10.1	14.2	10.9	12.4
14	74.6	76.7	50.8	55.3	5.6	6.0	5.9	8.0	7.4	7.6	10.5	10.1	9.8	10.4	7.4	7.8	15.0	16.9	15.0	13.5
15	71.8	100.0	51.4	69.5	6.2	7.6	8.8	8.8	7.3	10.3	9.4	14.1	11.0	14.2	6.4	8.0	14.5	22.7	15.6	17.1
16	93.3	95.9	63.7	64.5	6.8	6.8	9.6	9.0	7.9	9.1	12.7	14.5	12.1	13.2	8.3	7.9	20.0	21.0	19.3	18.2
17	95.3	98.3	65.3	67.0	7.2	7.1	9.4	8.8	9.1	9.3	12.7	14.2	12.0	13.7	9.3	8.6	20.9	21.1	19.1	18.6
18	95.0	99.7	61.4	66.7	6.9	7.7	10.0	9.0	9.3	10.2	13.8	15.2	11.9	13.3	8.3	8.8	17.2	19.8	20.5	18.6
19–21	92.8	109.6	59.6	71.3	7.0	7.4	10.2	9.0	8.1	10.1	14.0	18.7	11.7	15.2	7.9	8.8	16.4	22.9	19.4	21.2
22–24	82.0	102.6	54.0	68.7	6.4	7.1	9.3	9.9	8.3	9.5	11.5	15.6	11.5	14.3	8.0	8.7	12.0	21.1	16.6	19.6
25–29	98.4	83.5	60.1	52.1	6.2	5.7	10.6	8.3	9.0	7.8	16.1	12.3	13.3	11.1	8.0	7.2	14.5	13.7	23.2	18.7
30–34	83.1	90.7	53.9	55.7	6.4	5.5	9.7	8.6	8.2	8.9	13.0	15.0	10.7	11.9	7.9	7.5	13.6	14.8	20.8	19.4
35–39	75.8	91.4	47.1	55.6	5.4	5.5	9.1	8.3	6.8	7.8	12.1	15.8	8.9	13.3	6.3	7.3	11.3	12.6	17.1	21.9
40–44	93.7	90.9	54.7	53.7	6.3	6.1	9.6	8.7	7.7	8.6	16.4	16.8	12.1	11.6	8.0	7.4	11.9	13.2	21.7	21.2
45–49	78.7	82.1	46.6	50.7	5.4	5.8	9.0	8.2	6.1	7.4	14.1	13.4	9.1	11.4	6.7	6.5	12.1	11.1	19.0	18.2
50–54	67.8	97.2	38.9	57.1	5.4	6.5	8.3	8.7	6.4	8.3	11.1	18.7	8.3	14.1	5.9	7.8	7.8	14.7	18.9	23.6
55–59	72.5	84.4	43.8	47.2	5.9	6.1	9.5	8.5	6.3	8.1	11.3	16.4	11.1	10.7	5.3	6.7	8.3	8.9	20.0	21.8

* Score on total Alpha test, *minus* scores on tests 4 and 8.

TABLE III.—MEDIANS OF RAW SCORES OF FIVE HUNDRED EIGHTY-ONE MALES (M) AND SIX HUNDRED SEVEN FEMALES (F)

Age-group	Total alpha		Abbreviated alpha		Alpha subtest number																	
					1		2		3		4		5		6		7		8			
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
10	40.0	43.3	32.5	31.7	3.8	3.9	4.5	4.7	4.3	3.9	3.2	6.1	6.6	5.8	4.5	5.3	9.7	9.5	8.0	8.3		
11	40.0	52.5	30.0	34.4	4.8	4.9	5.7	5.2	3.9	4.0	4.1	5.9	6.4	7.8	5.4	5.3	7.0	8.6	8.1	6.5		
12	50.0	55.0	35.0	41.5	5.7	6.0	6.8	6.3	4.5	5.8	5.1	6.3	8.7	9.5	6.5	4.7	9.3	11.3	9.5	11.5		
13	50.8	66.3	37.5	49.2	5.4	6.6	7.5	7.4	5.6	6.2	4.8	7.7	8.6	10.2	5.5	6.9	9.8	14.0	9.7	11.9		
14	81.0	80.6	53.8	54.2	6.4	6.6	6.0	8.1	6.9	7.4	11.0	11.0	11.1	10.8	7.6	8.1	13.4	14.7	15.0	12.2		
15	71.3	96.0	52.1	68.8	6.8	8.2	9.0	8.6	7.4	10.0	9.8	12.6	11.9	13.0	6.4	9.0	13.0	23.7	14.7	16.1		
16	92.5	90.0	63.3	67.5	7.4	7.8	10.0	9.2	8.0	9.4	14.0	14.3	11.9	14.0	8.4	7.7	20.5	25.0	19.4	16.3		
17	92.5	92.5	66.3	67.5	7.3	7.6	9.0	9.2	8.2	9.3	13.0	12.7	12.0	13.8	9.2	8.8	20.0	23.0	19.3	14.5		
18	92.5	91.7	62.5	58.5	7.4	8.3	10.2	8.7	9.5	10.3	11.6	14.0	11.5	14.0	8.8	9.0	17.7	17.0	21.3	18.3		
19–21	75.0	101.7	48.6	72.1	7.8	7.8	9.5	8.8	7.4	9.8	10.8	17.0	11.3	15.0	7.7	9.3	13.0	24.2	17.0	20.5		
22–24	78.3	105.0	58.8	67.5	6.8	7.6	8.8	8.9	8.5	9.2	10.5	14.5	12.3	14.7	7.8	8.2	8.6	20.0	15.0	20.0		
25–29	93.0	83.6	54.2	51.1	6.6	5.8	10.5	8.1	8.8	7.5	15.0	12.6	12.7	11.3	8.6	7.5	12.0	11.5	24.2	19.7		
30–34	80.0	81.4	52.5	48.0	7.4	6.2	9.9	8.4	8.0	8.5	12.7	13.3	10.9	11.9	9.0	7.8	12.0	11.5	20.5	19.2		
35–39	65.0	88.3	41.7	53.2	5.6	6.0	9.1	8.4	6.3	8.2	11.0	15.0	8.0	13.8	6.8	7.7	7.7	11.3	14.0	22.1		
40–44	93.3	89.0	53.0	47.5	6.8	6.3	10.0	8.6	7.4	8.4	16.3	16.2	12.6	10.7	9.3	7.7	8.7	9.0	23.3	21.4		
45–49	67.0	80.8	40.8	48.2	6.2	6.3	9.3	7.7	5.6	7.5	14.6	14.2	9.5	12.2	7.6	6.3	9.8	9.8	17.0	19.3		
50–54	54.0	101.7	31.0	56.3	5.7	7.3	8.0	8.8	5.5	8.7	8.5	22.3	7.5	15.5	6.3	9.1	4.6	12.0	16.0	23.6		
55–59	60.0	81.7	40.0	42.5	6.0	6.5	8.6	8.4	5.7	8.0	7.0	17.4	10.5	11.5	2.7	6.7	7.0	7.8	17.0	21.4		

TABLE IV.—RELIABILITY OF THE DIFFERENCE BETWEEN FEMALE AND MALE PERFORMANCE ON THE ARMY ALPHA TEST*

Age-group	Total alpha		Abbreviated alpha		Alpha subtest number					
					1		2		3	
	Mean difference*	Reliability index*	Mean difference	Reliability index	Mean difference	Reliability index	Mean difference	Reliability index	Mean difference	Reliability index
10–15	9.8	5.3	7.5	6.1	.55	3.4	.29	1.6	.66	3.5
16–21	7.1	2.1	5.3	2.5	.28	1.4	−.88	−3.4	1.28	4.1
22–39	4.5	1.5	2.6	1.4	−.26	−1.4	−1.10	−4.5	.24	.9
40–59	9.1	2.6	5.5	2.6	.42	1.9	−.65	−2.0	1.38	4.8
10–59	8.3	6.2	5.8	6.8	.27	2.8	−.41	−3.5	.81	6.4

Age-group	Alpha subtest number									
	4		5		6		7		8	
	Mean difference	Reliability index	Mean difference	Reliability index	Mean difference	Reliability index	Mean difference	Reliability index	Mean difference	Reliability index
10–15	1.6	3.8	1.3	4.2	.61	2.6	3.0	5.5	.65	1.3
16–21	2.4	3.2	2.0	4.0	.10	.3	2.8	3.0	−.17	−.2
22–39	.8	1.1	1.1	2.5	.00	.0	1.5	2.0	−.17	−.2
40–59	2.3	2.6	1.8	3.4	.35	1.0	1.7	2.1	1.02	1.2
10–59	1.7	5.4	1.5	7.0	.33	2.2	2.4	6.6	.39	1.2

* In each case the mean score of the males has been subtracted from the mean score of the females. The derivation of the figures in the table may require a word of explanation. The basis of the present table is found in Tables I, II and V. For each of the eighteen age-groups in these tables, we obtained the mean difference (female *minus* male), and the PE of this difference by the usual formula (6, p. 133). Then, for the group including the ages ten, eleven, twelve, thirteen, fourteen, and fifteen, the mean difference at each age was weighted inversely as the square of its PE and the weighted mean difference for the group aged ten to fifteen was thus obtained (*cf.* 8, p. 325). (The weighted mean difference is given in the table under the heading "Mean difference.") A similar procedure was employed to obtain the weighted mean difference for the groups aged sixteen to twenty-one, twenty-two to thirty-nine, forty to fifty-nine, and ten to fifty-nine, respectively. The PE of the weighted mean difference was then obtained by the method recommended by Scarborough (10). The reliability index is the weighted mean difference divided by its PE. *A "positive" reliability index is here used to indicate female superiority; a negative index, female inferiority.*

Whatever their importance with reference to other problems, sampling factors do not disturb the comparison between age-sex differences in the various subtests. In four tests (tests 3, 4, 5 and 7) the males are rather consistently inferior. These are all strongly verbal tests. In test 1 (oral directions) the males are in early adolescence as inferior as in the more highly verbal tests; but at later ages they acquit themselves relatively much better, overtaking the females in the period of middle maturity, and again dropping slightly behind in later maturity. Test 8 (information) is one which has commonly been regarded as favoring males; the information items selected undoubtedly in many cases bear selectively upon male interests and occupations (11, pp. 227, 234). Our obtained sex differences are relatively slight in this test, and with frequent intersections of the growth curves. In test 6 (number series completions) a similar condition occurs—intersecting curves, with small differences, readily attributable to sampling fluctuations. But the similarity of the results for tests 6 and 8 can hardly be due to similarity of content. Test 6 is in fact more similar to test 2 (arithmetic problems). In this subtest we find the only instance in which males are fairly consistently superior to females. The male superiority does not clearly emerge until later adolescence, but thereafter is in striking contrast to the results for the verbal tests 3, 4, 5, and 7. This is probably an illustration of a sex difference which has been reinforced by differential practice and mental attitude (4). Tasks of computation fall traditionally to men, in connection with buying, trading, and planning works of farm construction. The farm wife describes herself as "no hand at figuring."[1]

It is clear, then, that by the manipulation of test material, striking changes could be produced in obtained sex differences. A heavier weighting of arithmetic elements would bring the male averages up to or above those for the females. A test made up largely of selected

[1] Possibly this division of tasks and responsibilities may be taken as a symptom of a patriarchal system surviving longer under rural than under urban conditions—the male of the household being the chief executive, secretary, and treasurer. In modern city life, the functions of budgeting and accounting are now shifting to an increasing degree to women. In considering these speculations we should not, however, neglect another hypothesis, which attempts to explain sex differences in arithmetical abilities in terms of inherent factors; women are naturally more proficient in the use of language; men excel in motor abilities, in handling objects, and in quantitative manipulation of all sorts. As yet, however, we are hardly prepared to do more than conjecture concerning the relative contributions of culture and of original nature in the development of these abilities.

items from subtests 3, 4, 5, and 7 would manifest an even more convincing female superiority, in the same group of subjects. Sex differences vary both in direction and degree according to the nature of the test. And whatever the role of native factors, it seems clear that differences in interest and practice—not themselves hereditarily determined—must also be considered. This is in agreement with the conclusion previously reached, by an analysis of sex differences on individual items of other tests administered to the same sample (4).

Taking the total Alpha for whatever it is supposed to measure, it must be conceded that the sex differences which we have shown are more largely of a theoretical than a practical significance. In early adolescence, the actual difference in Alpha score is only about one-third of an SD. This drops to one-sixth in later adolescence and to one-ninth in early maturity, increasing again to about one-fourth of an SD in later maturity.

Some interest attaches to sex differences in variability. These differences are hardly large enough to deserve statistical analysis, but in certain age groups they show a fairly marked consistency (Table V). In both sexes, marked increases in variability occur throughout the period of adolescence. This increase, however, appears to be greater for boys than for girls. After adolescence, the variability of males tends to hold constant, or to increase slightly in later maturity, while the female variability tends to diminish slightly. This latter would be in agreement with the assumption of a slightly superior selection of women after the age of forty-five.

In the range of ages from ten to sixty, the impressive fact is not the degree of sex difference, but rather the similarity of the developmental curves for the two sexes.

SUMMARY

The Army Alpha intelligence examination was administered to five hundred eighty-one males and six hundred seven females in a representative rural group between the ages of ten and sixty. A general slight superiority of females to males is observed; this superiority (as in physical growth curves) is greater during early adolescence, but is at no age sufficiently marked to be of great practical significance.

The sex difference is by no means uniform among the individual subtests of the Alpha. In four strongly verbal tests ("common sense," opposites, disarranged sentences, and analogies), the males are rather consistently inferior; in two tests (numerical completions and

TABLE V.—STANDARD DEVIATIONS OF RAW SCORES OF FIVE HUNDRED EIGHTY-ONE MALES (M), AND SIX HUNDRED SEVEN FEMALES (F)

Age-group	Total alpha		Abbreviated alpha		Alpha subject number															
					1		2		3		4		5		6		7		8	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
10	21.2	25.1	15.4	16.6	2.1	1.8	3.0	2.9	3.0	1.9	3.5	5.2	3.7	3.7	2.8	2.9	5.8	7.8	5.2	5.6
11	23.4	23.1	14.1	16.2	1.8	2.4	2.3	2.1	2.3	2.2	5.1	4.9	3.6	4.5	3.4	3.1	5.7	8.0	5.6	6.6
12	22.3	24.3	15.1	16.7	2.2	2.6	2.6	2.6	2.1	2.4	4.1	6.0	4.0	4.4	2.8	3.0	7.1	5.5	6.9	5.9
13	28.7	22.4	17.6	16.8	2.2	2.2	2.7	3.0	3.1	2.3	7.1	4.5	5.0	4.1	3.1	3.2	7.1	7.6	7.4	5.7
14	32.6	23.7	19.6	15.7	2.0	2.6	2.4	1.9	3.0	3.3	7.7	5.8	4.7	4.6	3.6	2.9	9.2	8.2	8.8	6.2
15	32.0	25.8	20.5	17.9	2.3	2.3	2.7	2.4	2.8	3.0	7.4	7.2	5.1	4.4	3.6	4.0	10.2	8.4	9.2	6.2
16	38.9	37.7	26.5	24.3	2.6	2.2	3.6	2.9	3.7	3.5	8.4	8.7	5.7	6.7	3.9	4.1	10.9	10.7	9.7	7.7
17	39.8	35.4	23.9	23.6	2.6	2.0	2.7	3.0	4.4	3.6	9.2	6.8	5.4	4.9	3.4	4.8	11.4	10.2	9.4	9.1
18	41.6	41.5	24.8	24.3	2.4	2.1	2.5	2.7	3.5	3.8	9.6	9.4	6.0	6.0	4.0	4.4	12.0	11.3	9.7	9.2
19–21	48.1	40.6	29.5	23.3	2.5	2.1	3.7	2.6	3.7	3.8	10.7	10.3	6.8	6.1	4.9	4.2	12.9	10.6	10.4	8.3
22–24	39.0	41.7	22.7	25.4	2.3	1.9	3.0	2.8	4.5	3.4	9.2	9.9	5.1	6.4	4.2	4.6	10.0	12.3	10.4	8.5
25–29	42.4	33.5	24.4	22.1	2.7	2.4	2.9	3.4	3.6	3.2	9.7	8.3	5.8	5.4	4.1	3.9	10.7	8.8	9.7	8.0
30–34	43.8	43.9	25.9	26.7	2.7	2.7	3.6	3.4	4.1	3.8	10.3	10.9	6.7	6.1	3.9	3.9	10.6	11.6	10.1	10.1
35–39	44.1	40.1	25.4	22.4	2.9	2.6	3.7	3.2	3.7	2.9	10.8	10.5	6.3	6.5	3.9	4.1	9.5	8.9	10.7	8.7
40–44	45.2	44.3	25.9	27.8	3.0	3.1	4.1	4.1	4.0	3.4	10.6	10.5	6.9	7.1	4.4	4.3	10.3	11.2	10.1	9.5
45–49	44.1	35.7	25.3	22.7	2.5	2.3	3.3	3.0	3.0	3.0	10.4	9.3	6.4	5.3	3.8	4.0	10.1	9.5	10.4	8.7
50–54	43.0	37.4	26.0	24.9	2.3	2.2	4.7	3.8	4.1	2.9	10.3	10.8	6.4	6.3	4.5	4.0	9.4	10.5	10.9	7.7
55–59	47.5	30.6	25.3	17.3	2.7	1.8	4.2	2.8	3.9	2.6	12.0	8.8	6.7	5.4	5.0	4.1	7.3	7.3	11.7	6.8

general information), the sex differences are relatively slight, with frequent intersections of the developmental curves; in one test (arithmetic problems), the males are quite definitely superior.

These results confirm conclusions reached by an item analysis of other tests administered in the same sample. It is clear, then, that the direction and extent of sex differences in a composite mental test is dependent upon the composition of the test and the weighting of its parts. Further investigation of "intelligence" differences between the two sexes should be made in terms of specific functions or items, rather than in terms of unanalyzed total scores.

REFERENCES

1. Carter, H. D.: Family Resemblances in Verbal and Numerical Abilities. *Genetic Psychology Monograph*, Vol. XII, 1932, pp. 1–104.
2. Conrad, H. S.: General-information, Intelligence and the Decline of Intelligence. *Journal of Applied Psychology*, Vol. XIV, 1930, pp. 593–599.
3. Conrad, H. S. and H. E. Jones: Psychological Studies of Motion Pictures. III. Fidelity of Report as a Measure of Adult Intelligence. *University of California Publications in Psychology*, Vol. III, 1929, pp. 245–276.
4. ——: Psychological Studies of Motion Pictures. V. Adolescent and Adult Sex Differences in Immediate and Delayed Recall. *Journal of Social Psychology*, Vol. II, 1931, pp. 433–459.
5. ——: A Field Study of the Differential Birth Rate. *Journal of the American Statistical Association*, Vol. XXVII, 1932, pp. 153–159.
6. Garrett, H. E.: "Statistics in Psychology and Education." New York: Longmans Green, 1926.
7. Jones, H. E. and H. S. Conrad: The Growth and Decline of Intelligence. A Study of a Homogeneous Group between the Ages of Ten and Sixty. *Genetic Psychology Monograph*, Vol. III, 1933.
8. Kelley, T. L.: "Statistical Method." New York: Macmillan, 1924.
9. Price, B.: "The Perceptual Ability of Persons over Fifty Years of Age." M. A. Thesis, Stanford University, 1931.
10. Scarborough, J. B.: The Invalidity of a Commonly Used Method for Computing a Certain Probable Error. *Proceedings of the National Academy of Sciences*, Vol. XV, 1929, pp. 665–668.
11. Yerkes, R. M.: Psychological Examining in the United States Army. "Memoirs of the National Academy of Sciences," Vol. XV. Washington: Government Printing Office, 1921.

THE COLLECTING INTERESTS OF TOWN CHILDREN AND COUNTRY CHILDREN

PAUL A. WITTY AND HARVEY C. LEHMAN

Northwestern University and Ohio University

In two articles which appeared recently in this JOURNAL[1] the writers discussed certain phases of children's interest in collecting. In the first of the articles, data regarding the extent to which town children report interest in collections were presented. In addition, the problem of terminology was discussed. The fact was emphasized that "collection" is a popular term which lacks precise scientific meaning, and that when attempts are made to employ this term scientifically, vagueness and ambiguity often result. In the second article, salient sex differences in collecting interests and activities were presented with the writers' interpretation thereof. This article deals with another aspect of collecting; data will be presented comparing the collecting interests of town and country children.

In *Whither Mankind*,[2] C. A. Beard and others have ably discussed certain of the changes in man's mode of living that have been precipitated by the advent of the machine age. This book, and most others of its kind, barely touches upon the changes in child life that have been brought about by the industrial revolution. This neglect is due, doubtless to several facts; among these the lack of pertinent data is perhaps the most important. Nevertheless, a moment's reflection concerning the enormous social and economic changes which have accompanied the industrial upheaval will bring to mind many conditions that must inevitably be reflected in child life. For example, less than a century ago, in 1860, only three per cent of the people of the United States lived in cities. At the present time, approximately half of the population of the United States resides in cities. Such an unprecedented shift in the habitat of the population has brought numerous problems of adjustment for children and for adults. In his introductory speech before the White House Conference, President Hoover set forth the following opinions.

> In the last half century we have herded 50,000,000 human beings into towns and cities where the whole setting is new to the race. We have created highly congested areas with a thousand changes resulting in the swift transition from a rural and agrarian people to an urban, industrial nation. Perhaps the widest

range of difficulties with which we are dealing in the betterment of children grows out of this crowding into cities. Problems of sanitation and public health loom in every direction. Delinquency increases with congestion. Overcrowding produces disease and contagion. *The child's natural play place is taken from him.* His mind is stunted by the lack of imaginative surroundings and lack of contact with the fields, streams, trees and birds.*,[3] (Italics not in the original.) * P. 274.

At a recent conference of educators, Norton described some values which the typical farm child of a generation ago gained from his out-of-school life.

He performed the chores with which the farm abounds. As soon as he was old enough, he aided in such fundamental farm operations as sowing, hoeing and harvesting. When the horses needed shoeing he drove with his father to the blacksmith shop and saw the glowing iron fashioned at the forge and fitted to the horse's foot. The milk bottle did not mark the limit of his knowledge as to the source of milk. He did not have to go to a zoo to see a cow. He knew all about their habits, and the chores connected with them from first-hand experience. Nearly every hour of the day brought him increased understanding of, and training for, life. He knew how real work was done, and learned how to do it.

Compare this situation with that of the child living in the congested city. There is much less in the typical urban community that children can do. The grocer brings the eggs, machines milk the cow, and the baker bakes and delivers the bread. Actual participation in the work of the city is prohibited by child labor laws; and it is difficult to observe what goes on even at second hand, for "keep out" is posted in large letters at the entrance to the factory and shop. Often turned loose with nothing to do, children have to choose between play upon the crowded city street, association with the back alley hoodlum, or attendance at the cheap movie show.*,[3] * P. 274.

With increased urbanization of America's population, it becomes increasingly important for the educator to know as much as possible about the modern child's needs. Although it is impossible to ascertain in detail the differences in the environments of children who lived forty or fifty years ago and of those today, it is possible to discern significant differences in the environments of present-day town and country children. These data should be *suggestive* of differences which have been brought about (in part at least) by urbanization. The environmental differences of present-day town and country children may be set in sharp relief by behavior studies, one type of which is concerned with the child's collecting activities. Although the present study does not include data for children living in congested city areas, it does present comparative data for children living in a small city and for those living in a genuinely rural environment.

METHOD OF INVESTIGATION

A check-list of one hundred ninety items was presented to large numbers of town and country children. The procedure that was used for assembling the one hundred ninety items and the directions which were given to the pupils may be ascertained by reference to a former article.[1(a)] The town children lived in Lawrence, Kansas, a university center having a population of approximately 15,000. Data for rural children were obtained from the rural schools in Medina County, Ohio.

RESULTS

Table I presents the average number of active collections reported by rural children of various school grades. Comparison of these data with those presented previously for town children makes it

TABLE I.—AVERAGE NUMBER OF ACTIVE COLLECTIONS REPORTED BY COUNTRY CHILDREN OF VARIOUS SCHOOL GRADES

School grade	Number of boys	Average number of collections	Number of girls	Average number of collections
2	81	8.22	54	9.67
3	118	13.54	127	19.73
4	149	13.85	118	19.43
5	157	19.74	143	23.43
6	127	10.69	116	12.48
7	108	11.99	101	15.43
8	72	9.41	57	9.26
9	67	5.42	54	8.20
10	63	6.11	50	12.52
11	29	6.52	35	7.43
12	37	6.51	27	6.22
Total.....	1008	.....	882	

evident at once that the rural child collects *many more objects* than does the city child. This fact is revealed by several lines of evidence, some of which will receive later comment.

It will be noted from Table I that the largest average number of active collections is found for rural children in the fifth grade. In studies of town children it has been found rather consistently that the largest average number of collections is made by children of age ten. This finding suggests that the country child matures somewhat more slowly than does the city child; typical rural children sometimes are found to be slightly older than city children for their respective grades.

Table II presents items which were collected two or more times as frequently by country boys as by town boys. For the convenience of the reader the items of Table II have been classified in ten categories. It will be noted at once from Table II that the frequency distributions for the first three categories (botanical specimens, animal parts and insects, and rocks and minerals), give clear evidence that the rural child has closer contact (and doubtless better acquaintance) than does the town child with objects of nature. This fact is of course one that would naturally have been anticipated.

Certain of the items of Table II, particularly those listed under the heading, "Miscellaneous," will doubtless appear to the reader to be of a rather trivial nature. Many of these resemble possessions more than collections. This fact might lead one to conclude that the great frequency of reported collections among the rural children may be due to misunderstanding of the term "collection." The thought may occur to the reader that the rural child reported not only his collections, but also provided an inventory of his possessions. Nevertheless, several analyses of the data indicate that the rural child actually collects much more frequently than does the town child. For example, Table II presents seventy-eight items which were collected two or more times as often by country boys as by town boys.* Although seventy-eight items were collected two or more times as often by country boys as by towns boys, only one item (gopher skins†) was collected more than twice as often by the town boys as by the country boys. The data for the girls are similar to those of the boys in this respect. (See Table III.) Of the items included in the check-list, one hundred nineteen were collected two or more times as often by country girls as by town girls. Only two items (jack-stones and toys [great variety]) were collected two or more times as frequently

* Five items of Table II were collected slightly less than twice as often by country boys as by town boys. These five items were included in Table II because they seemed to fit well into the general categories of this table.

† This doubtless is due to the fact that gophers are rare in Ohio.

TABLE II.—ITEMS COLLECTED TWO OR MORE TIMES AS OFTEN BY COUNTRY BOYS AS BY TOWN BOYS

	1008 Country boys, per cent	808 Town boys, per cent
(a) Botanical		
Acorns	7	3
Clover (four-leaf)	15	6
Ferns	5	1
Grass (kinds of)	3	1
Moss (kinds of)	3	1
Rose petals	2	1
Oak balls	3	1
(b) Animal parts, insects, etc.		
Bees	6	2
Birds' eggs	8	4
Birds' beaks and claws	3	1
Birds' nests	8	3
Cats (different kinds)	4	1
Cocoons	9	3
Dragon flies	2	1
Snails	5	2
Skeletons	2	1
Skins	6	3
Teeth	5	1
Rabbit feet	4	0
(c) Rocks and minerals		
Crystals	5	2
Flint	17	6
Fossils	6	3
Minerals	4	2
Shells	12	6
Soils (samples of)	3	1
Stones (kinds of)	12	5
Sand, colored	4	1
(d) Salable items		
Iron	13	6
Rags	6	2
Rubber	7	3
Tinfoil	18	2
Junk to be sold	16	1
Milk bottles to be sold	5	0
(e) Objects associated with hunting, fishing, hiking, etc.		
Arrowheads	27	6
Bullets or cartridges	17	9
Fish hooks	14	9
Guns (toy)	8	5
Knives	14	9
Shot	9	4
Wampum	2	1

TABLE II.—*Continued*

	1008 Country boys, per cent	808 Town boys, per cent
(*f*) Objects for personal adornment		
Belts	7	3
Buttons, cuff	4	2
Jewelry	6	3
Neckties (No. of kinds)	9	4
Scarf pins	2	0
Stick pins	2	0
Rings	9	3
(*g*) Objects which possess sentimental value		
Cards (election, merit)	8	4
Letters	8	3
Charm strings	2	1
Spoons	3	1
Monograms	1	0
(*h*) Souvenirs from the schoolroom		
School books	16	6
Drawings	11	6
Compositions	3	1
(*i*) Objects essential to certain types of play and recreation		
Song books	6	3
Blocks	4	2
Dogs (toy)	4	1
Pants guards	2	1
(*j*) Miscellaneous		
Bells	4	2
Booklets	9	3
Flags	10	4
Glasses	3	1
Horse shoes	15	8
Matches	5	2
Match box tops	5	2
Newspaper covers	4	1
Tissue paper	3	1
Seals	4	2
Tin tags	3	1
Tobacco tags	3	1
Paint (kinds of)	4	2
Boxes	16	6
Clocks	7	3
Dishes (broken)	3	1
Sacks (kinds of)	4	2
Paint charts	3	1
Calendars	7	3
Thread	2	1
Strings	8	2
Seeds (kinds of)	8	4
Shoes (old ones)	4	1
Old hats	4	0

TABLE III.—ITEMS COLLECTED TWO OR MORE TIMES AS OFTEN BY COUNTRY GIRLS AS BY TOWN GIRLS

	882 Country girls, per cent	868 Town girls, per cent
(*a*) Botanical		
Acorns	6	2
Beans (kinds of)	6	2
Ferns	11	4
Clover (four-leaf)	20	13
Pressed flowers	22	11
Grass (kinds of)	5	1
Moss (kinds of)	6	3
Nut shells (kinds of)	6	2
Oak balls	3	1
Wood (samples of)	4	2
Rose petals	7	4
Trees (samples of)	5	2
(*b*) Animal parts, insects, etc.		
Bees	4	1
Beetles	2	1
Birds' eggs	5	2
Birds' beaks and claws	1	0
Birds' nests	8	2
Birds' wings	2	1
Butterflies	11	5
Cocoons	8	2
Dragon flies	2	1
Furs	5	2
Rabbit ears	2	1
Snails	5	1
Skins	2	0
Spiders	1	0
Teeth	4	2
Rabbit feet	2	0
(*c*) Rocks and minerals		
Crystals	5	2
Flint	8	1
Fossils	6	2
Minerals	3	1
Sand, colored	4	1
Shells	13	6
Stones (kinds of)	16	5
Soil (samples of)	3	1
(*d*) Salable items		
Bones	3	1
Brass	3	1
Iron	5	1
Junk to be sold	9	0
Lead	6	2
Metals (kinds of)	5	2
Rags	5	2
Rubber	4	1
Tin	2	1
Tinfoil	20	2
Milk bottles to be sold	4	0
(*e*) Objects associated with hunting, fishing, hiking, etc.		
Arrowheads	10	1
Bullets or cartridges	2	0
Fish hooks	4	2
Guns (toy)	2	1
Knives	4	2
Shot	2	0
Wampum	1	0

TABLE III.—*Continued*

	882 Country girls, per cent	868 Town girls, per cent
(*f*) Objects for personal adornment		
Badges	5	2
Beads	24	12
Belts	8	2
Bracelets	14	7
Breast pins	9	3
Buttons	11	4
Buttons, pictures	5	1
Buttons, cuff	4	1
Neckties (No. of kinds)	4	1
Pins, hat	7	1
Pins, scarf	6	1
Pins, stick	4	1
(*g*) Souvenirs from the schoolroom		
Drawings	21	13
Compositions	8	4
School books	27	12
(*h*) Objects essential to certain types of play and recreation		
Blocks	5	1
Doll buggies	8	3
Dogs (toy)	4	2
Kites	5	2
Marbles	16	5
Marbles (agate)	9	3
Tops	4	2
Story books	27	3
Song books (kinds of)	19	9
(*i*) Miscellaneous		
Glasses (pieces of)	6	3
Blank bank checks	5	2
Bells	5	1
Cigar papers	4	2
Cigar tins	2	1
Glasses (for drinking)	6	2
Horse shoes	6	3
Labels	6	1
Matches	3	1
Match box tops	6	2
Monograms	3	1
Newspaper covers	3	1
Sayings (all descriptions)	8	3
Pants guards	2	0
Tags, shipping	2	0
Time tables	6	2
Tobacco sacks	2	1
Trade marks	3	1
Wagon wheels	2	1
Dishes (broken)	8	4
Bottle caps	7	3
Boxes	14	6
Coins	20	10
Cigar tags	2	1
Jugs	5	1
Nails	5	2
Needles	8	3
Oil cans	2	0
Padlocks	3	1
Paint (kinds of)	4	1
Paint charts	4	2
Tissue paper	9	4
Pieces of quilts	10	4
Pins, common	8	3
Rubber sheets	2	0
Sacks (kinds of)	3	1
Old shoes	5	2
Strings	8	4
Thread	8	4
Vases	14	4
Old hats	6	0
Clothes-pins	6	2
Clocks	4	2
Cups and saucers (kinds of)	9	3
China (painted)	9	4

by town girls as by country girls. Very few items were collected more often by town children than by country children. (See Tables IV and V.) The data appear to present rather conclusive evidence of the relatively great interest of the country child in collecting.

TABLE IV.—ITEMS COLLECTED SOMEWHAT MORE OFTEN BY TOWN BOYS THAN BY COUNTRY BOYS

	808 Town boys, per cent	1008 Country boys, per cent
Bottles	8	7
Cigar tags	4	3
Cigar stamps	6	4
Curios	4	3
*Gopher skins	4	1
Marbles	30	22
Newspapers	17	9
Poems	5	4
*Rabbit ears	6	4
Stamps	18	13
Wood, petrified	4	3

* Both gophers and rabbits are much more common in Kansas than in Ohio.

The fact should perhaps be mentioned that among the rural children, the sex differences in collecting interests were found to be almost identical with those which the writers have reported previously

TABLE V.—ITEMS COLLECTED SOMEWHAT MORE OFTEN BY TOWN GIRLS THAN BY COUNTRY GIRLS

	868 Town girls, per cent	882 Country girls, per cent
Autograph sentiments	14	13
Jackstones	9	4
Newspapers	11	9
Toys (great variety)	4	1

for town children.[1(b)] Instances in which a previously reported sex difference was reversed are few indeed. The validity of the general categories employed previously for describing sex differences in collecting interests are corroborated therefore by the present findings.

In a study that was reported in 1900, C. F. Burk wrote as follows:

> It will be seen that the boys concentrate more on a few things which run as crazes through very many of them . . . nearly half of them had collected birds' eggs. Nothing attains such widespread interest among the girls, although between thirty and forty per cent of them had collected stamps, shells and picture cards, but a greater number of things have considerable runs with the girls. The table shows that only seven things were collected by ten or more per cent of the boys—while sixteen things were collected by ten or more per cent of the girls.[4] P. 186.

Tables VI and VII show that the writers' studies fail to corroborate Burk's finding that "the boys concentrate more on a few things which seem to run as crazes." Nevertheless, these data do support Burks' finding that "a greater number of things have considerable runs with

TABLE VI.—EXTENT TO WHICH TOWN AND COUNTRY CHILDREN "CONCENTRATE ON A FEW THINGS WHICH SEEM TO RUN AS CRAZES"

	Town children		Country children	
	Boys	Girls	Boys	Girls
Largest per cent collecting any one item.....	30	28	27	28
Number of items collected by ten or more per cent of individuals.....	6	23	30	55
Number of items collected by twenty or more per cent of individuals.....	2	2	4	11

the girls." But the foregoing finding is probably due merely to the fact that girls exceed boys in the average number of collections made. In Table VI it will be noted that thirty items were collected by ten or more per cent of the country boys and that only six items were collected by ten or more per cent of the town boys. One might conclude from this that "a greater number of things have considerable runs with" the country boys. This conclusion would be unwarranted and extraneous because the average number of collections of rural boys exceeds that of city boys.

Burk's report that "the boys concentrate more on a few things which seem to run as crazes," was due probably to local or to temporary conditions. Miss Whitley failed to corroborate this finding*,[5] and the

* P. 254.

writers' data also do not substantiate it. When the present data are partitioned by school grades, one sees that, in nine out of eleven grades,

TABLE VII.—LARGEST PERCENTAGE OF COUNTRY BOYS AND COUNTRY GIRLS COLLECTING ANY ONE ITEM

School grade	Most popular item	Boys, per cent	Most popular item	Girls, per cent
2	Marbles (also tinfoil)	29	Story books (also hdkfs)	40
3	Marbles	33	Valentines	34
4	Marbles	26	Valentines	34
5	Arrowheads	35	Story books	40
6	Arrowheads	31	Valentines	27
7	Marbles	31	Valentines	34
8	Arrowheads	28	Drawings (different kinds)	28
9	Arrowheads	28	Souvenirs	35
10	Arrowheads	25	Souvenirs	48
11	Arrowheads	24	Souvenirs (also photographs)	40
12	Arrowheads	26	Dance programs	40

TABLE VIII.—ITEMS COLLECTED MORE OFTEN BY TOWN BOYS THAN BY TOWN GIRLS, BUT MORE OFTEN BY COUNTRY GIRLS THAN BY TOWN BOYS

	868 Town girls	808 Town boys	882 Country girls	1008 Country boys
Snails	1	2	5	5
Birds' eggs	2	4	5	8
Birds' nests	2	3	8	8
Jugs	1	3	5	4
Rubber	1	3	4	7
Tobacco tags	0	1	2	3
Arrowheads	1	6	10	27
Flint	1	6	8	17
Story books	3	9	27	16
Pants guards	0	1	2	2
Rubber sheets	0	1	2	2
Sacks (kinds of)	1	2	3	4
Tags (shipping)	0	1	2	2
Trade marks	1	2	3	3
Matches	1	2	3	5

the country girls "concentrated more" than did the country boys in collecting certain objects. (See Table VII.)

The fact is perhaps worthy of mention that certain items which were collected more often by town boys than by town girls were nevertheless more frequently collected by country girls than by town boys. (See Table VIII.) Similarly, certain items which were collected by town girls more often than by town boys were collected more often by country boys than by town girls. (See Table IX.) Such findings suggest that environmental influence is in some instances more potent than is sex in determining the extent of collecting interests.

TABLE IX.—ITEMS COLLECTED MORE OFTEN BY TOWN GIRLS THAN BY TOWN BOYS, BUT MORE OFTEN BY COUNTRY BOYS THAN BY TOWN GIRLS

	868 Town girls	808 Town boys	862 Country girls	1008 Country boys
Ferns	4	1	11	5
Grass (kinds of)	1	1	5	3
Hat pins	1	0	7	2
Stick pins	1	0	4	2
Scarf pins	1	0	6	2
Clover (four-leaf)	13	6	20	15
Strings	4	2	8	8
Old shoes	2	1	5	4
School books	12	6	27	16

It is difficult not to speculate regarding these data. For example, why should the country child collect so much more frequently than the town child? Numerous factors of course are operative. It is the belief of the writers that every child has to work out an adjustment to his environment by trial-and-error experience. Since it is often difficult for the country child to find congenial playmates of his own age, he is obliged often to engage in individual endeavor. Collecting is an activity that is easily accessible and apparently congenial for the country child. And the rural child avails himself of this form of amusement with unusual frequency.

Nevertheless, many other factors are doubtless operative in stimulating the urge to collect. Possibly in his collecting, the rural child is simply imitating to a goodly extent the adult life that forms a part of his social environment. There is frequently space on the farm for worn-out or partially used articles of various sorts to be stored. Since the junk-man rarely makes his rounds in the country such articles do accumulate. A tendency toward habitual thrift may also be a

causal factor of no little import in producing the country child's excessive interest in collecting. Furthermore, in some instances the child may be merely following the line of least resistance, since it may be easier for him to allow objects to accumulate than to destroy them.

PEDAGOGICAL IMPLICATIONS

Collecting is frequently an individualistic, competitive type of behavior. The child rarely engages in collecting as a member of a group or of a team. He often tells about or displays "my" collection rather than "our" collection. He collects frequently because he sees other children collecting and his goal is likely to be that of obtaining a larger or a more valuable collection than his rival.

What effect this individualistic, competitive type of behavior is likely to have upon numerous attitudes of the rural child is a matter of conjecture. It seems plausible, however, in the light of the laws of habit-formation, that the collecting endeavor of the rural child might make of him an individualistic adult. Today many assert that the American farmer is inclined to be individualistic in his outlook and in his attitudes—too individualistic for his own good. The charge is often brought that farmers are incapable of coöperating politically (and also in other ways). Whether this charge be valid or not, the data in this paper suggest that geographical isolation might well be a factor that makes for individualism.

A foremost objective of the modern school is expressed by the single word, "socialization." This being the case, it seems plausible that, if interest in collections is a fair sample of his generalized behavior, the rural child is in even greater need of a socialized school program than is the town child. The frontiersman, the pioneer, and the farmer have doubtless lived in environments that made for individualistic outlook. Their children have also been subjected largely to similar influences. But if the modern child is to live his adult life in an urbanized environment, it is necessary that he learn coöperation.

In other places the writers have pointed to the differences between the play activities of town and of country children.[6] Some of the differences are probably to the advantage of the town child; some of them probably favor the rural child. Country children were found to engage more often than town children in play activities such as the following:

Blackman (Pom-pom-pullaway)	Visiting or entertaining company
Horse-back riding	Teasing somebody
Hunting	Throwing rocks or stones
Shooting	Swinging
Andy-over	Wading in the water
Teeter-totter	Climbing porches, trees, fences, posts, etc.
Gathering nuts	Using a hammer, saw, nails, etc., for fun
Hop, skip, and jump	Cutting paper things with a scissors
Looking at pictures	Drawing with pencil, pen, chalk, or crayon
Just singing	Playing with pet dogs, pet kittens, etc.
Whistling	

Town children were found to engage more often than country children in the following:

Football	Going to the movies
Basket-ball	Listening to the radio
Riding a bicycle	Doing gymnasium work
Playing cowboy	Watching athletic sports
Boxing	Going to parties or picnics
Marbles	Looking at the Sunday "funny paper"
Roller-skating	Playing robber and police
Playing Indian	

Very interesting in this connection is the fact that town life tends to suppress certain natural modes of self-expression, for example, whistling and singing. Whistling and singing were found by the writers to be participated in by country children much more commonly than by town children. Although rural life seems to permit freer expression to certain natural modes of behavior, it affords less opportunity for participation in certain organized group activities.* To the foregoing statements the following may be added, *viz.*, rural life permits the child to maintain his contact with the objects of nature that are listed in the first three categories of Tables II and III. Rural life tends also to direct the child to individual endeavor. One of the outcomes of the latter fact is a marked tendency on the part of the country child to make collections.

REFERENCES

1. (*a*) Witty, P. A., and Lehman, H. C.: Further Studies of Children's Interest in Collecting. *Journal of Educational Psychology*, Vol. XXI, 1930, pp. 112–127.

* The writers have found also that, as compared with boys, girls have relatively few team games in which they may participate. This condition probably obliges girls to engage more often in individual endeavor and probably accounts in part for the fact that girls collect more frequently than do boys.

(*b*) ——: Sex Differences: Collecting Interests. *Journal of Educational Psychology*, Vol. XXII, 1931, pp. 221–228.

2. Beard, C. A. (Editor): "Whither Mankind." Longmans, Green and Co., New York, 1928, pp. vii + 408.
3. Norton, J. K.: Education and the Economic Depression. *School and Society*, Vol. XXXIV, 1931, pp. 271–277.
4. Burk, C. F.: The Collecting Instinct. *The Pedagogical Seminary*, Vol. VII, 1900, pp. 179–207.
5. Whitley, M. T.: Children's Interest in Collecting. *Journal of Educational Psychology*, Vol. XX, 1929, pp. 249–261.
6. Lehman, H. C., and P. A. Witty: "The Psychology of Play Activities. A. S. Barnes & Co., New York, 1927, pp. xviii + 242.

DISTRIBUTIONS OF SCORES BASED ON CORRECT GUESSING FOR TRUE-FALSE TESTS OF VARIOUS LENGTHS

WILLIAM C. F. KRUEGER

College of the City of Detroit, Detroit, Michigan

The purpose of this study was (1) to determine experimentally the distributions of frequencies based on the number of correct guesses in "true-false tests" of various lengths; (2) to find a practical length which would eliminate the probability of getting a high score by chance guessing; (3) to observe the function of the formula, "Right minus Wrong equals True Score," when applied to scores obtained from tests of various lengths.

The subject received a sheet of paper containing either 10, 20, 30, 40, 50, 60, 80, 100, 150, 200, 300, or 500 serially numbered lines. He was asked to imagine that he was writing a "true-false test." Each line on the page was to represent a statement which he was to judge as "true" or "false." Since the line could not furnish a cue with respect to the truth or falsity of the test item, the subject was forced to guess. (The straight line was chosen instead of the worded statement because the worded statement might suggest an answer to the subject.) In some of the tests the subject was asked to mark in random order fifty per cent of the statements "true" and the rest "false." He was instructed not to use any scheme in marking, for example, he was not to alternate the "true" and "false" answers. For the other tests he was requested to mark the answers at random as "true" or "false" without keeping count of the frequency of the "true" or "false" signs. Again no scheme in answering was to be followed.

All papers were scored by a key in which fifty per cent of the answers, chosen at random, were "true" and the rest "false." The frequency distributions of the number of items guessed correctly as "true" or "false" were tabulated in Tables I and II. Every frequency distribution for each length of test is based on the judgments of one thousand subjects. The class intervals in Table I are in units of ten per cent, such as one per cent to ten per cent, eleven per cent to twenty per cent, twenty-one per cent to thirty per cent, etc. The class intervals in Table II are in units of five per cent. The second classification contains data from Table I, but it shows more accurately the distribution of scores, particularly for the longer series. Since

TABLE I.—DISTRIBUTIONS OF SCORES BASED ON CORRECT GUESSES FOR THE TRUE-FALSE TESTS RANGING FROM 10 TO 500 ITEMS PER TEST

I, per cent	Number of items in the test											
	10 items		20 items		30 items		40 items		50 items		60 items	
	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2
0	.....	1										
1–10	12	9										
11–20	44	37	12	7	4	2	2					
21–30	101	95	90	70	49	37	22	17	12	3	5	2
31–40	208	230	230	205	143	141	164	146	120	96	103	76
41–50	260	254	330	302	398	403	401	407	434	486	488	504
51–60	201	210	232	250	296	320	318	334	358	350	360	366
61–70	103	102	96	133	101	90	87	90	74	61	44	52
71–80	48	43	8	29	8	6	6	6	2	3		
81–90	15	13	2	4	1	1						
91–100	4	6										
SD.....	16.38	15.72	11.90	12.61	10.44	9.83	9.49	9.15	8.39	7.65	7.39	7.11
PE.....	.35	.33	.25	.27	.22	.21	.20	.19	.18	.16	.16	.15

I, per cent	80 items		100 items		150 items		200 items		300 items		500 items	
	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2
0												
1–10												
11–20												
21–30												
31–40	59	54	31	21	14	3	6	4				
41–50	501	492	507	531	543	556	574	555	515	522	522	524
51–60	406	412	438	433	435	437	420	440	485	477	478	476
61–70	34	42	24	15	8	4	....	1	....	1		
71–80												
81–90												
91–100												
SD.....	6.55	6.62	5.98	5.65	5.39	5.11	5.05	5.06	4.97	5.02	4.99	4.99
PE.....	.14	.14	.13	.12	.11	.11	.11	.11	.11	.11	.11	.11

Explanations: B^1—Frequencies for the respective class intervals when the subject limited his guesses to 50 per cent "True" and 50 per cent "False."

C^2—Frequencies for the respective class intervals when the subject guessed at random and did not mark one-half of the items "True" and the other "False" by actual count.

I—Class intervals in steps of 10 per cent of the total number of possible answers. The "0 per cent" class interval records the frequencies of 0 items guessed correctly.

one person guessed zero per cent correctly in the ten-items tests, Table I also contains a special class interval for zero per cent guessed correctly. The corresponding SD's and PE's are also given underneath the respective columns.

TABLE II.—DISTRIBUTIONS OF SCORES BASED ON CORRECT GUESSES FOR THE TRUE-FALSE TESTS RANGING FROM 20 TO 500 ITEMS PER TEST

I	Number of items in the test															
	20 items		40 items		60 items		80 items		100 items		200 items		300 items		500 items	
	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2	B^1	C^2
1–5																
6–10																
11–15	3	2														
16–20	9	5	2													
21–25	21	31	5	4	1											
26–30	69	39	17	13	4	2										
31–35	101	84	69	50	27	22	12	15	5	2						
36–40	129	121	95	96	76	54	47	39	26	19	6	4				
41–45	147	137	168	171	183	201	182	153	137	164	92	110	28	54	8	12
46–50	183	165	233	236	305	303	319	339	370	367	482	445	487	468	514	512
51–55	131	142	192	186	201	215	235	242	298	256	360	358	449	431	464	465
56–60	101	108	126	148	159	151	171	170	140	177	60	82	36	46	14	11
61–65	61	74	64	61	38	44	29	33	22	12	....	1	....	1		
66–70	35	59	23	29	6	8	5	9	2	3						
71–75	6	23	5	4												
76–80	2	6	1	2												
81–85	1	2														
86–90	1	2														
91–95																
96–100																
SD.....	11.67	12.39	9.26	9.00	7.15	6.89	6.32	6.35	5.44	5.42	3.79	4.05	3.07	3.36	2.72	2.72
PE.....	.25	.26	.20	.19	.15	.15	.13	.13	.12	.12	.08	.09	.07	.07	.06	.06

Explanations: I—Class intervals in steps of five per cent of the total number of answers possible.

B^1—Frequencies of guesses for each respective class interval when the subject was instructed to mark fifty per cent of his guesses "True" and fifty per cent "False."

C^2—Frequencies of guesses for each respective class interval when the subject was directed to guess at random without an attempt to mark as many times "True" as "False."

It can readily be observed that in both tabulations there is a definite trend which indicates that the longer the test is, the greater is the frequency of correct guesses within the range from forty-one per cent to sixty per cent of the number of items contained in the test.

If one applies the formula, "Right minus Wrong equals the True Score," to the various scores obtained, many unusually high and low

(negative) scores are obtained from the ten-items and twenty-items tests. However, in the longer tests, practically every score yields a final mark of approximately zero.

If length of a test is plotted against the Standard Deviation for the scores obtained from that test, the resultant curves are negatively accelerated and indicate a consistent relationship between the two factors.

CONCLUSIONS

1. A definite and obvious trend indicates that the longer the test is, the greater is the frequency of scores within the class intervals ranging from forty-one per cent to sixty per cent of the number of items in the test. For the longer tests, practically all scores ranged within forty-five per cent and fifty-five per cent of the total test.

2. Chance guessing may frequently yield very high and very low scores in short tests. In the longer tests this probability is practically eliminated.

3. When the formula, "Right minus Wrong is the True Score," is applied to the longer series, it furnishes a fairly adequate means to eliminate the probability of getting high scores by guessing only.

TEACHING OBJECTIVES IN EDUCATIONAL PSYCHOLOGY

ROBERT A. DAVIS

University of Colorado

Educational psychology has served as a blanket term to include the fields of measurement, statistics, child psychology, mental hygiene and learning. Colleges and universities offer specific courses for treating the subject matter listed under the general term educational psychology; consequently students taking courses so designated often duplicate previous work. The topic of learning receives little attention, whether it is presented in a separate course or as one phase of a general course in educational psychology. This tendency may be due to the inadequacy of research and to the fact that available studies emphasize nonsense syllables, sorting of cards, rote memory and motor skills and do not stress those types which are characteristic of learning in the classroom. An analysis[1] of objective studies in educational psychology yields evidence that research in learning is ill-adapted to the teacher's use and that results of investigations are conflicting and therefore confusing to the teacher of educational psychology.

Although there are obvious limitations to research in learning the major emphasis in a course in educational psychology should be placed upon learning and those internal and external factors which bear upon its development. The courses in educational psychology developed for students of education should especially stress learning. If it is agreed that educational psychology should direct attention to the problems of learning there is the further question of selecting the content for a course in this field. In addition to those factors which underlie learning ability there are a number of internal and external influences which should be considered as topics in their relation to the ease and permanency of learning. A partial solution to the problem of content may be found by an examination and classification of objective studies which deal with learning in its various forms under

[1] Davis, Robert A. and C. R. Ballard: The Development of Research in Learning. *Journal of Educational Psychology*, February, 1932.

See also Henmon, V. A. C.: Needed Research in Learning. *Journal of Educational Research*, Vol. XI, 1925, pp. 313–321.

diverse conditions. When such studies are classified according to established criteria there is found a large number of not clearly differentiated classes of subject matter in educational psychology. However, the functions of these materials as they are related to the field and student have not been clearly defined and outlined, and there is no agreement as to objectives[1] among textbooks in educational psychology. As a result there is neither conformity in objectives nor subject matter in this field. Objectives should be established with two purposes in mind; one set of objectives should be outlined from the standpoint of the field of educational psychology itself, and the other with the view to making clear the important abilities and skills which the subject should serve to develop in students.

In establishing objectives in educational psychology definite assumptions must be made. It must be assumed that the student has been prepared in basic courses, and in those topics usually covered in general psychology. In addition he should possess some of the pertinent facts dealing with physical and mental growth, and emotional and social reactions as they apply to childhood and youth. In general the student should have some appreciation of the intricate inner life of the individual which would create insight and sympathy for the growing mind. These are minimum requirements and considerably more training in the techniques of experimental research and psychology would be needed for those students of advanced standing who intend to carry forward work for higher degrees in education or psychology. For advanced students the objectives would necessarily differ from those outlined for prospective high-school teachers who expect educational psychology to contribute toward their professional training. The subject matter of educational psychology is so inextricably allied with functions and objectives, that no attempt is made to differentiate between outcomes realized by the students and the content employed. Educational psychology should serve:

1. *To Develop an Understanding and Appreciation of Those Hereditary and Environmental Factors Which Underlie Learning Ability.*— Investigations dealing with the influences of heredity and environment should be studied and evaluated, the emphasis being placed upon mental rather than biological traits. This objective also includes a consideration of the nature of improvement, and an evaluation of

---

[1] Good, Warren R.: The Need for Objectives in Educational Psychology. *School of Education Bulletin*, University of Michigan, April, 1931.

the theories proposed for its explanation; the inter-relationship of mental processes, the nature and significance of intelligence, and the character of learning ability. A foundation and point of view may be established which will provide the structure upon which later objectives may be based.

2. *To Provide Bases for Understanding the Nature and Principles of Learning and to Supply the Techniques for Guidance in Learning.*—An interpretation of physiological and psychological learning should be considered with the theories of learning, and the laws operating to affect economy in the learning process. In addition the student should have drill in the application of these learning principles to class-room conditions. Educational psychology should develop ability in the selection and application of techniques which will elicit the most favorable learning responses. This application of learning principles should be adapted to the varying degrees of maturity and ability of pupils so that there is a progressive adjustment of materials and techniques to pupils as they develop through the several stages of mental and educational growth.

3. *To Provide a Foundation for General and Specific Methods of Teaching.*—This objective can also be attained through an understanding of learning principles and laws. Methods of teaching are valid only when based upon a knowledge of the ways by which learning takes place and those things which influence it. Teaching techniques consist chiefly in presentation of materials and provision of incentives for the attainment of definite ends. The function of methodology is to develop techniques for making effective the operation of the laws of learning. Although each school subject is probably individual enough to justify separate treatment in a course or topic, the general laws of learning are sufficiently comprehensive to serve as guiding criteria for all learning situations.

4. *To Train the Student in the Recognition of the Different Types of Learning and to Develop Ability in the Application of These Types to the Various Fields of Subject-matter.*—Since the principles of learning do not apply with equal force to all types of learning, educational psychology should develop the ability to differentiate between those types of learning which involve predominantly motor responses and those which evoke mental reactions from pupils. Again, training should be directed to the differentiating process and to applying the types of learning to the learning and teaching of school subjects. As between motor and mental learning which are not clearly differentiated, the

major emphasis should be given to mental learning of the sort which is employed by pupils in mastering the abstract subjects of the school. Attention should also be directed to the technique of reasoning and reflective thinking. Training in the technique of reflective thinking is one of the aims of all education but the appropriate place for its development is in the field of educational psychology.

5. *To Train the Student in the Recognition and Significance of Individual Processes and Traits.*—Under this objective is noted the part played by such factors as drives, the development of perception, observation, interest, and attention. Other factors are the influence on ability and achievement of physical traits and various types of sensory defects. The bearing of these factors on learning and on the adjustment of the child to the school are basic in the study of educational psychology and should be correlated with the general theme of learning. The aim is to recognize the existence and development of those traits and mental processes which are concomitants of child development as they relate to learning and teaching. Although these factors are susceptible to training and education, they are primarily a phase of native equipment, and may, therefore, be considered as predominantly inherent in character.

6. *To Provide Understanding and Ability in the Recognition of Those External Factors Which Are Largely within the Control of the Teacher and School.*—In one case the factors bearing on learning are predominantly inherent in the individual and in the other they are largely external in nature and within the control of the teacher and school. These factors include modes of presentation as related to reception of stimuli, the use of guidance techniques in teaching, the influence of continuous work and fatigue, schedules, motivating devices, ventilation, and humidity of classrooms. These influences have been studied with varying purposes in mind both in the laboratory and school but the important conclusions of investigators may be related to the problem of learning efficiency. From the standpoint of the function of the school improvement is affected largely by the development of techniques of studying and learning and by the provision of favorable school environment.

7. *To Provide the Student with Ability and Skill in the Methods of Measuring and Evaluating Teaching Efficiency.*—Although the student may take measurement courses, educational psychology should train in measurement as it relates to learning and teaching. Educational psychology should not only develop techniques for producing learning

but it should also serve to provide ability and skill in measuring the effectiveness of teaching devices and methods. Measurement is essential in educational psychology and is used in all investigations dealing with learning. However, this objective should emphasize the significance and use of standardized measuring instruments for estimating the amount and quality of pupil achievement for definite periods of instruction. Through the use of objective measuring instruments, the teacher can test the relative effectiveness of methods and techniques of teaching which have been employed. This training should deal with the concepts of reliability, validity, and objectivity as they apply to measures of achievement in the various school subjects. The efficiency of teaching and the degree of pupil achievement can be reliably determined only by the use of objective examinations.

GENERAL OBJECTIVES

In addition to the foregoing teaching objectives, there are general objectives which should be observed at the outset of study in educational psychology. These objectives include those of information, attitudes, skill, and appreciation which in a degree are common to all fields of education. Educational psychology should serve:

1. *To Provide the Student with a Body of Factual Data and Methods Which Can Be Used in Solving His Own Teaching Problems.*—From the viewpoint of the use of objective tools and devices educational psychology is a pioneer among the fields of education, and experimental methods of research have been the chief means by which problems in this field have been studied. As a result a considerable body of scientific material was available before the other fields of education had begun to use the techniques of objective research. The student should derive a body of factual information from his study and possess some knowledge of the methods and techniques used by investigators for arriving at the findings from which his information is obtained. Too frequently information is secured but no understanding of the means by which it is produced and interpreted. Some knowledge of the techniques of experimental research is essential for an appreciation of the readings in educational psychology.

2. *To Develop the Scientific and Problem-solving Attitude of Students.* Due to the controversial nature of the problems in the field of educational psychology and the refinement of tools and techniques used in the study of these problems the conclusions of research studies are constantly modified. It is important that the student begin his

study of educational psychology with the determination not to accept any conclusion as final for which conclusive proofs are not available. In no other way can he acquire an understanding of the method and spirit of science in education and of educational psychology in particular.

3. *To Train the Student to Think Psychologically about Educational Problems.*—Educational psychology seeks to apply psychological principles to the problems of learning and teaching. These psychological principles apply in one case to the pupil who is the learner, and in the other to the teacher who organizes materials of instruction, and stimulates desires so that learning may be effectively guided. The student should therefore, think in psychological terms of pupils, teachers and methods of teaching as they are functionally and interdependently related.

REFERENCES

Davis, Robert A.: Changing Tendencies in General Psychology. *Psychological Review*, Vol. XXXVI, No. 4, July, 1929, pp. 320–331.

Davis, Robert A. and C. R. Ballard: The Development of Research in Learning. *Journal of Educational Psychology*, December, 1932.

Henmon, V. A. C.: Needed Research in the Field of Learning. *Journal of Educational Psychology*, Vol. XI, 1925, pp. 313–321.

Hertzberg, O. E.: The Opinions of a Teacher-Training Institution Concerning the Relative Value of Subject-matter in Educational Psychology. *Journal of Educational Psychology*, May, 1928.

Watson, G. B.: What Shall Be Taught in Educational Psychology? *Journal of Educational Psychology*, December, 1926.

Good, Warren R.: The Need for Objectives in Educational Psychology. *School of Education Bulletin*, University of Michigan, April, 1931.

A SIMPLIFIED STATEMENT OF THE TWO-FACTOR THEORY

W. LINE AND H. B. HEDMAN[1]

University of Toronto

1. PURPOSE OF THIS PRESENTATION

Students of psychology are, quite properly, brought up in an atmosphere of controversy. From their earliest acquaintance with the subject, varying points of view are contrasted and compared, so that the subtleties of mental science are increasingly appreciated. Scientific doubt is fostered; new systematic positions are examined critically, and attacked from many angles; so that he who would express his own individuality by formulating an original system of psychological concepts, must be prepared to defend himself against a vast array of spirited rivals, and to convince sceptics of a most obstinate kind.

This state of affairs is a healthy one, in a field as complex as that of human behaviour, where personal opinion is so vitally interested, and hence where temperamental bias can be overcome only with great difficulty. At the same time, however, it is quite obvious that, if carried too far, doubt becomes a handicap; for one's appreciation of a system may be seriously impaired by excessive caution, such as frequently occurs when the system involves the acceptance of the conclusions of allied sciences. The field of knowledge has become so vast, that scholarship of the comprehensive type manifested in a Newton, or attempted in a Spencer, is no longer possible. Hence the study of any one branch of science involves a combination of faith and scepticism. Doubt carried to the Cartesian extreme is likely to constitute a negative influence, to become an end in itself, and, in the realm of psychology, to be other than wholly impersonal. None the less, when the major roots of a system penetrate the heart of another field, the ambitious student is to some extent justified in demanding at least a logical demonstration of the validity of the primary postulates.

Spearman's "Two-Factor Theory" is a case in point. In order to establish a scientific methodology that would raise certain controver-

[1] The mathematical portion of this paper is mainly the work of the second author.

sies in the Mental Test field above the level of mere opinion, a new approach to the problem of validity was demonstrated in 1904.[1] This approach established the view that the tendency to rely upon a purely *empirical* interpretation of validity,—as followed, for example, by those who would be content with the fact that "intelligence tests" *work* in practice,—merely represented a recourse to obscurantism, or something worse. At the same time, any attempt to go beyond this empirical stage towards the truly psychological problem as to the nature of the function measured, (the "*functional*" validity,) was doomed to failure, since the controversies thus engendered could never be overcome without an adequate methodology. The truly "*objective*" validity could not, prior to that time, be determined. Hence the logical approach to the study of mental constitution through the field of Individual Differences, involved the discovery and demonstration of a method that would determine the various "abilities" or "factors" in a rigidly objective manner. The functional and empirical aspects would be later considerations.

This means that, fundamental to a scientific approach in the field of psychology is an understanding of the techniques designed to measure degrees of correspondence between functions.[2] The mathematics of correlation must be understood and developed to meet the needs of psychological science. Galton implied the same thing when he said that a better understanding of the inner constitution of mind should be possible;[3] and Spearman has given concrete expression to Galton's idea, and has shown that upon a rigidly objective foundation, a scientific account of functional psychology can be built.[4] This latter task involves all the intricacies of the mental realm; but the methodology makes them amenable to objective validation.

The psychologist would be justified in assuming that the validity of the mathematical techniques involved, was not his immediate concern. Just as we accept the formulae of the natural scientist and apply them without necessarily understanding their derivation, so we might employ statistical devices in evaluating our data without

---

[1] Spearman, C.: The Measurement of Association between Two Things and General Intelligence—Objectively Determined and Reasured: *American Journal of Psychology*, Vol. XV, 1904.

[2] Spearman, C.: *Loc. cit.*, p. 71.

[3] Inquiries into Human Faculty, (Everyman's Library, 1919 Edition, p. 17).

[4] The functional aspect is treated fully in Spearman's "The Nature of Intelligence," etc. (Macmillan, 1923), where the doctrine of Neogenesis is presented.

concerning ourselves unduly with the reasoning on which they are based, especially when such devices have been checked by the leading experts in the field of mathematics. We must, of course, understand them sufficiently to be able to apply them without error. But beyond that, we may undoubtedly accept the substantiation of a Filon,[1] or of a Wilson.[2]

Nevertheless, since the mathematical aspect plays so important a part in the epoch-making work of Professor Spearman and his school, the student may be pardoned for a feeling of discomfort if asked to build on a foundation that is, to him, unknown. With this in view, the material which follows has been prepared. It constitutes the first steps of an attempt to present the basic theorems underlying the factor theory, in a manner that can be appreciated by those unacquainted with the higher mathematics. To the expert, there may be certain unavoidable limitations in such a procedure; but, it is hoped, no false steps in the reasoning are involved. The application of the formulae to typical psychological problems has not been considered. Such an extension of the *Appendix* to Spearman's *Abilities of Man* has already been made available by Holzinger.[3] An amplification and simplification of the main sections of the *Appendix* is here presented, to assist in a more comprehensive understanding of the factor theory and of its far-reaching possibilities. For those who are interested in the more technical presentations of these theorems, and of the controversies that at one time surrounded them, a working bibliography is appended.

2. PLAN OF THE WORK

The significance of the sequence here adopted is indicated by the following outline:

Section A

(*a*) Introductory statement concerning the meaning of "Correlation."

(*b*) The development of a formula for determining the coefficient of correlation.

---

[1] See, for example, *Proceedings of Royal Societies of London* 1922, Vol. 101 A, pp. 94–100.

[2] Wilson, E. B.: *Science*, March 2, 1923.

[3] Holzinger, K. J.: Statistical Résumé of the Spearman Two-factor Theory, U. of Chicago Press, 1930.

(*c*) The effects of errors upon the coefficient of correlation, (r).
(*d*) Hence the need for "correlation" of r.
(*e*) The Correction for Attenuation.
(*f*) The Tetrad Criterion, (as derived from *e*).
(*g*) The significance of the Tetrad Criterion.
(*h*) The Reliability of the Tetrad Criterion.
(*i*) Freedom of the Tetrad Criterion from the influence of Attenuation.

Section B

(*a*) Alternative method of correcting for Attenuation.
(*b*) Alternative method of deriving the Tetrad Criterion;—Partial Correlation.

Section C

Symbols used in this paper.

Section D

Bibliography for technical study of the Two-factor Theory.

SECTION A

(*a*) *The Meaning of Correlation as Applied to Mental Abilities.*—The term "correlation" refers to the degree of inter-dependence between two mental functions. Expressed quantitatively, perfect correlation is signified by a coefficient of 1, perfect independence by 0. For inverse correlation the coefficients take similar but negative values. On this basis it is possible to express the relationships between functions or abilities in simple numerical terms. The main relationships met in the field of cognitive abilities are positive, and range from 1 to 0.

The most frequently used expression for determining the coefficient of correlation (r) between two functions, is that known as the Pearson Product-Moment formula:

$$r = \frac{\Sigma xy}{n\sigma_x\sigma_y} \qquad \text{(I)}$$

The immediate interpretation of this formula may be obtained by consulting the table of Symbols employed throughout this paper. (Section C, page 218.) We are here concerned with a logical interpretation of Equation (I), which is best arrived at by showing how it is derived.

(*b*) *Development of a Formula for Determining the Coefficient of Correlation.*—Suppose we wish to determine the degree of correlation

between Ability A and Ability B, (*e.g.* between "arithmetical" and "mechanical" abilities), as measured in a sample of Grade 8 children. Suppose further, that the scores obtained in tests of A and B are indicated by the distributions x and y respectively. We may plot the scores obtained by the individuals in both tests on the same diagram, by representing the x variable horizontally, and the y variable vertically, and thus construct a scatter diagram, or a correlation table.

Y-variable (Mechanical test scores →)									
									1
					1	2	3	1
				1	2	4	2	1
			1	2	6	5	3	
		1	2	4	5	3	1	
		2	3	4	2	1		
	1	2	3	2	1			
	2	1						
1								

Arithmetical test scores →
X-variable

Before comparing the x and y distributions, we must reduce them to comparable units; *i.e.*, we must change the point scores to their values in terms of the equivalent amounts of deviation from their means. This is done by dividing the deviations (in points) from the mean of the distribution by the standard deviation. Thus an x deviation in the "arithmetical" distribution equals x/σ_x in terms of its standard unit; and a y deviation in the "mechanical" distribution equals

y/σ_y in terms of the corresponding unit. We may now express the ratio between any corresponding pair of deviations.

Let r_1 denote the ratio between any x deviation in a row and its corresponding y deviation, each having been reduced to sigma terms. It is obvious that the y value for a row is constant.

Let n denote the number of x deviations in a row.

For a single x deviation in a row:

$$\frac{x}{\sigma_x} = r_1 \frac{y}{\sigma_y}$$

or

$$x = r_1 \frac{\sigma_x}{\sigma_y} y \qquad \text{(i)}$$

Summing for a row:

$$\Sigma_x = \Sigma r_1 \frac{\sigma_x}{\sigma_y} y$$

$$\frac{\Sigma x}{n} = \frac{\Sigma r_1}{n} \frac{\sigma_x}{\sigma_y} y$$

But

$$\frac{\Sigma_x}{n} = \text{the mean } x \text{ of the row,}$$

and

$$\frac{\Sigma r_1}{n} = \text{the ratio of the mean } x \text{ of a row to the } y \text{ of that row, which value will now be called } r, \text{ since this is the coefficient required.}$$

Therefore, in any row:

$$\text{Mean} \qquad x = r \frac{\sigma_x}{\sigma_r} y \qquad \text{(ii)}$$

This equation measures the regression of x on y.

Returning to equation (i), for a single deviation in a row

$$x = r_1 \frac{\sigma_x}{\sigma_y} y$$

$$xy = r_1 \frac{\sigma_x}{\sigma_y} y^2$$

For a row

$$\Sigma xy = \Sigma r_1 \frac{\sigma_x}{\sigma_y} y^2$$

$$\Sigma xy = n \frac{\Sigma r_1}{n} \frac{\sigma_x}{\sigma_y} y^2$$

$$\Sigma xy = r \frac{\sigma_x}{\sigma_y} n y^2$$

since $\Sigma r_1 / n = r$.

For the whole table:

$$\Sigma(\Sigma xy) = r \frac{\sigma_x}{\sigma_y} \Sigma(ny^2),$$

since r, the slope of the regression line x on y, is constant for the whole table.

But obviously

$$\Sigma(\Sigma xy) = \Sigma xy$$

and

$$\Sigma(ny^2) = \Sigma y^2.$$

$$\therefore \Sigma xy = r \frac{\sigma_x}{\sigma_y} \Sigma y^2$$

$$= r \frac{\sigma_x}{\sigma_y} n \cdot \sigma_y^2$$

Since $\sigma_y = \sqrt{\Sigma y^2 / n}.$

$$= r \cdot n \sigma_x \sigma_y.$$

$$\therefore r = \frac{\Sigma xy}{n \sigma_x \sigma_y} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} \qquad \text{(iii)}$$

This expression gives the coefficient of correlation in terms of x and y deviations.

In the above process, we passed from the equation for a row to an equation for the entire table, by the process of summation, and assumed that the means of the x scores of the rows lie approximately on a straight line, (the regression line); we also assumed that r, being the slope of that line, is constant for all the rows. By a similar process we can find the regression of y on x:

Mean $$y = r \frac{\sigma_y}{\sigma_x} x \qquad \text{(iv)}$$

and from this equation we may deduce as before, that $r = \Sigma xy / n \sigma_x \sigma_y$.

The important function of the regression equation is that of prediction. Given a child's "mechanical" score, we are able, by means of equation (iv), to predict his probable "arithmetical" score, *i.e.* the average score in arithmetic corresponding to the given mechanical score. It should be noted that the coefficient of correlation in itself has no predictive value.[1]

(*c*) *The Effects of Errors upon the Coefficient of Correlation.*—We may examine the influence of errors upon both numerator and denominator of the formula for r derived above:

We saw that $r = \Sigma xy/n\sigma_x\sigma_y$ for true values of the variables involved. Let the observed values be represented by

$$r_1 = \frac{\Sigma x_1 y_1}{n\sigma_{x_1}\sigma_{y_1}}$$

Consider first the effect of errors on the numerator of the coefficient (Σxy):

$$x_1 = x \pm e_{x_1}$$
$$y_1 = y \pm e_{y_1}$$
$$\therefore x_1 y_1 = xy \pm xe_{y_1} \pm ye_{x_1} + e_{x_1}e_{y_1}$$
$$\therefore \Sigma x_1 y_1 = \Sigma xy \pm \Sigma xe_{y_1} \pm \Sigma ye_{x_1} + \Sigma e_{x_1}e_{y_1}$$

If we assume that there is no correlation between either x or y and the errors, and that there is no correlation between the errors themselves, then

$$\sqrt{x \,.\, e_{y_1}} = \frac{\Sigma xe_{y_1}}{n\sigma_x\sigma e_{y_1}} = 0$$
$$\therefore \Sigma xe_{y_1} = 0$$

Similarly

$$\Sigma ye_{x_1} = 0$$

and

$$\Sigma e_{x_1}e_{y_1} = 0$$
$$\therefore \Sigma x_1 y_1 = \Sigma xy.$$

Thus we see that the numerator of the coefficient remains unchanged.

---

[1] The simple proof that the two regression lines pass through the mean of the distributions has been omitted in the above development. This proof may be found in any good text-book on Statistics. Cf. Yule, Udny: "An Introduction to the Theory of Statistics," p. 170.

Turning to the effect of errors on the denominator ($n\sigma_x\sigma_y$) of the coefficient, we arrive at a quite different result:

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n}}$$

and

$$\sigma_x{}^2 = \frac{\Sigma x^2}{n}$$

$$\sigma_{x_1} = \sqrt{\frac{\Sigma x_1{}^2}{n}}$$

and

$$\sigma_{x_1}{}^2 = \frac{\Sigma x_1{}^2}{n}$$

$$x_1 = x \pm e_{x_1}$$
$$x^2{}_1 = x^2 \pm 2xe_{x_1} + e^2{}_{x_1}$$
$$\Sigma x^2{}_1 = \Sigma x^2 \pm 2\Sigma xe_{x_1}\ \Sigma e^2{}_{x_1}$$

Applying the same assumption as before, regarding the correlation between x or y and the errors, and between the errors themselves,

$$\Sigma x^2{}_1 = \Sigma x^2 + \Sigma e^2{}_{x_1}$$
$$\therefore \frac{\Sigma x^2{}_1}{n} = \frac{\Sigma x^2}{n} + \frac{\Sigma e^2{}_{x_1}}{n}$$
$$\therefore \sigma^2{}_{x_1} = \sigma^2{}_x + \sigma^2{}_{ex1}$$
$$\therefore \sigma^2{}_{x_1} > \sigma^2{}_x \text{ and hence } \sigma_{x_1} > \sigma_x.$$

Similarly $\sigma_{y_1} > \sigma_y$, so that $n\sigma_{x_1}\sigma_{y_1} > n\sigma_x\sigma_y$.

The denominator is thus *increased* by errors, while the numerator remains unchanged. Therefore $\Sigma x_1y_1/n\sigma_{x_1}\sigma_{y_1}$ is always less than $\Sigma xy/n\sigma_x\sigma_y$, and consequently r_1 is always less than r.

(*d*) *The Need for Correction of r.*—From the foregoing it is obvious that errors of observation always have an attenuating influence upon the coefficient of correlation. This being so, the need arises for some device whereby the extent of that influence may be determined, or else whereby the errors may be eliminated, since our major objective is the accurate determination of degrees of correspondence between abilities. To this task we may now proceed.

(*e*) *The Correction for Attenuation.*—Let us begin with the expression for r in terms of *true* measures, and convert it into an expression in terms of *observed* measures. We have:

$$r_{xy} = \frac{\Sigma xy}{n\sigma_x\sigma_y} \text{ (true measures)}$$

and

$$r_{x_1y_1} = \frac{\Sigma x_1y_1}{n\sigma_{x_1}\sigma_{y_1}} \text{ (observed measures).}$$

Required: r_{xy} in terms of observed measures.

I. ANALYSIS OF THE NUMERATOR (Σxy)

We have already seen that $\Sigma xy = \Sigma x_1y_1$ (page 202). By exactly similar reasoning, we may determine the value of Σxy in terms of other sets of measures of the same abilities, *e.g.* x_2 and y_2. Thus x_1 and x_2 may denote the measures for two tests of discrimination, y_1 and y_2 those for two tests of intelligence, while x and y represent the *true* measures. So that, expressing Σxy in terms of the various combinations of measures, we have:

$$\left.\begin{aligned}\Sigma xy &= \Sigma x_1y_1\\ \Sigma xy &= \Sigma x_2y_2\end{aligned}\right\} \therefore (\Sigma xy)^2 = \Sigma x_1y_1 \cdot \Sigma x_2y_2. \quad \text{(i)}$$

$$\left.\begin{aligned}\Sigma xy &= \Sigma x_1y_2\\ \Sigma xy &= \Sigma x_2y_1\end{aligned}\right\} \therefore (\Sigma xy)^2 = \Sigma x_1y_2 \cdot \Sigma x_2y_1. \quad \text{(ii)}$$

II. ANALYSIS OF THE DENOMINATOR ($n\sigma_x\sigma_y$)

$$n\sigma_x\sigma_y = n\sqrt{\frac{\Sigma x^2}{n}}\sqrt{\frac{\Sigma y^2}{n}} = \sqrt{\Sigma x^2 \Sigma y^2}$$

As before:

$$x_1 = x \pm e_{x_1}$$
$$x_2 = x \pm e_{x_2}$$
$$\therefore x_1x_2 = x^2 \pm xe_{x_2} \pm xe_{x_1} + e_{x_1}e_{x_2}$$
$$\therefore \Sigma x_1x_2 = \Sigma x^2 \pm \Sigma xe_{x_2} \pm \Sigma xe_{x_1} + \Sigma e_{x_1}e_{x_2}$$

Again applying the assumption that the errors are uncorrelated, and that they are independent of x and y, (page 202)

$$\Sigma xe_{x_2} = 0$$

and

$$\Sigma xe_{x_1} = 0$$

and

$$\Sigma e_{x_1}e_{x_2} = 0$$
$$\therefore \Sigma x_1x_2 = \Sigma x^2$$

In a similar way we find the value of Σy^2 in terms of y_1 and y_2, and obtain the result:

$$\Sigma y_1 y_2 = \Sigma y^2.$$
$$\therefore \Sigma x^2 \Sigma y^2 = \Sigma x_1 x_2 \cdot \Sigma y_1 y_2$$
$$\therefore \sqrt{\Sigma x^2 \Sigma y^2} = \sqrt{\Sigma x_1 x_2 \Sigma y_1 y_2}$$
$$\therefore \text{The Denominator, } n\sigma_x \sigma_y = \sqrt{\Sigma x_1 x_2 \Sigma y_1 y_2}$$

We may now express the value of r_{xy} in terms of x_1, x_2, y_1, y_2 thus:

$$r_{xy} = \frac{\sqrt{\Sigma x_1 y_1 \Sigma x_2 y_2}}{\sqrt{\Sigma x_1 x_2 \Sigma y_1 y_2}} \qquad \text{(i)}$$

or

$$r_{xy} = \frac{\sqrt{\Sigma x_1 y_2 \Sigma x_2 y_1}}{\sqrt{\Sigma x_1 x_2 \cdot \Sigma y_1 y_2}} \qquad \text{(ii)}$$

But

$$r_{x_1 y_1} = \frac{\Sigma x_1 y_1}{n\sigma_{x_1} \sigma_{y_1}} = \frac{\Sigma x_1 y_1}{\sqrt{\Sigma x_1^2 \Sigma y_1^2}},$$

whence

$$\Sigma x_1 y_1 = r_{x_1 y_1} \cdot \sqrt{\Sigma x_1^2 \cdot \Sigma y_1^2}$$

And similarly

$$r x_2 y_2 = \frac{\Sigma x_2 y_2}{\sqrt{\Sigma x_2^2 \Sigma y_2^2}}, \; r_{x_1 x_2} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2 \Sigma x_2^2}}, \; r_{y_1 y_2} = \frac{\Sigma y_1 y_2}{\sqrt{\Sigma y_1^2 \Sigma y_2^2}}.$$

Substituting these values in equation (i), we get:

$$r_{xy} = \frac{\sqrt{(r_{x_1 y_1} \cdot \sqrt{\Sigma x_1^2 \cdot \Sigma y_1^2})(r_{x_2 y_2} \sqrt{\Sigma x_2^2 \Sigma y_2^2})}}{\sqrt{(r_{x_1 x_2} \cdot \sqrt{\Sigma x_1^2 \Sigma x_2^2})(r_{y_1 y_2} \cdot \sqrt{\Sigma y_1^2 \Sigma y_2^2})}}$$
$$= \frac{\sqrt{r_{x_1 y_1} \cdot r_{x_2 y_2}}}{\sqrt{r_{x_1 x_2} \cdot r_{y_1 y_2}}} \qquad \text{(i)}$$

And similarly, from equation (ii),

$$r_{xy} = \frac{\sqrt{r_{x_1 y_2} \cdot r_{x_2 y_1}}}{\sqrt{r_{x_1 x_2} \cdot r_{y_1 y_2}}} \qquad \text{(ii)}$$

Dividing (i) by (ii) we obtain the equation

$$1 = \frac{\sqrt{r_{x_1y_1} \cdot r_{x_2y_2}}}{\sqrt{r_{x_1y_2} r_{x_2y_1}}} = \frac{r_{x_1y_1} \cdot r_{x_2y_2}}{r_{x_1y_2} \cdot r_{x_2y_1}}$$

$$\therefore r_{x_1y_1} \cdot r_{x_2y_2} = r_{x_1y_2} \cdot r_{x_2y_1}.$$

This is the *Tetrad Equation.*

Multiplying (i) by (ii) we get

$$r^2_{xy} = \frac{\sqrt{r_{x_1y_1} \cdot r_{x_2y_2} \cdot r_{x_1y_2} r_{x_2y_1}}}{r_{x_1x_2} r_{y_1y_2}}$$

$$\therefore r_{xy} = \frac{(r_{x_1y_1} \cdot r_{x_2y_2} \cdot r_{x_1y_2} \cdot r_{x_2y_1})^{1/4}}{(r_{x_1x_2} \cdot r_{y_1y_2})^{1/2}}$$

This is the formula for the *Correction for Attenuation*, given by Spearman in *The Abilities of Man.* (Appendix, I.)

(*f*) *The Tetrad Criterion.*—In the preceding section, it was noted that the Tetrad Equation forms an integral part of the reasoning on which the correction for attenuation is based. Later on, (Section B) we shall consider the development of this equation from a different angle. Our task at this point is to state more clearly what the tetrad criterion is, and thence we shall pass to a brief consideration of its significance (Sub-section *g*).

In the formulae given on page 202, x_1 and x_2 stood for two sets of measures of function x, y_1 and y_2 for two sets of measures of function y. It was shown that

$$r_{x_1y_1} \cdot r_{x_2y_2} = r_{x_1y_2} \cdot r_{x_2y_1}$$

This expression would be true if x and y were correlated perfectly; *i.e.* if they represented identical functions; for then, by the "attenuation" formula, we should have

$$r_{xy} = 1 = \frac{(r_{x_1y_1} r_{x_2y_2} r_{x_1y_2} r_{x_2y_1})^{1/4}}{(r_{x_1x_2} r_{y_1y_2})^{1/2}}. \qquad (a)$$

Hence

$$(r_{x_1y_1} r_{x_2y_2} r_{x_1y_2} r_{x_2y_1})^{1/4} = (r_{x_1x_2} r_{y_1y_2})^{1/2}$$

and since, by hypothesis, $x = y$, y_1 and y_2 are measures of x. Consequently we may interchange x_2 and y_1, and get

$$(r_{x_1x_2} r_{y_1y_2} r_{x_1y_2} r_{x_2y_1})^{1/4} = (r_{x_1y_1} r_{x_2y_2})^{1/2} \qquad (b)$$

Similarly we may interchange x_2 and y_2 and get

$$(r_{x_1y_1} r_{x_2y_2} r_{x_1x_2} r_{y_1y_2})^{1/4} = (r_{x_1y_2} r_{y_1x_2})^{1/2} \qquad (c)$$

Dividing b by c,

$$\frac{(r_{x_1y_1}r_{x_2y_2})^{\frac{1}{2}}}{(r_{x_1y_2}r_{y_2x_2})^{\frac{1}{2}}} = \frac{(r_{x_1x_2}r_{y_1y_2}r_{x_1y_2}r_{x_2y_1})^{\frac{1}{4}}}{(r_{x_1y_1}r_{x_2y_2}r_{x_1x_2}r_{y_1y_2})^{\frac{1}{4}}}$$

When simplified, this becomes

$$r_{x_1y_1}r_{x_2y_2} = r_{x_1y_2}r_{y_1x_2}. \quad \text{(Appendix I, 2.)}$$

An experimental demonstration of such functional identity was given by Spearman in 1904,[1] in which the tests used were of many and varied types in the realm of cognitive abilities. Using his main result in illustration of the above presentation, we may take our symbols as standing for the following:

x_1 = discrimination of musical pitch.
x_2 = discrimination of visual length.
y_1 = power to succeed in school work.
y_2 = intelligence as measured by tests.

It can be assumed that x_1 and x_2 are measures of some ability x common to both discrimination functions—the measures involving, however, certain errors e_{x_1} and e_{x_2} respectively. Similarly y_1 and y_2 may be said to measure a common function y, with errors e_{y_1} and e_{y_2}. The correlation r_{xy}, corrected for attenuation, may thus be determined; and Spearman's results proved this to be practically perfect. On this basis, therefore, there is ground for the hypothesis employed above in illustrating the meaning of the Tetrad equation. [Appendix I, 1.]

(*g*) *The Significance of the Tetrad Criterion.*—The foregoing indicates that, when each of four variables, $(x_1x_2y_1y_2)$ can be divided into two factors each, the one being common to all four, (x or y), whilst the other is in each case specific and independent, (e_{x_1}, e_{x_2}, e_{y_1}, e_{y_2}), the tetrad equation obtains. This has now been investigated over a very wide range of abilities, and its empirical demonstration may be said to be beyond question. The reverse problem, namely, as to whether, when the tetrad criterion is satisfied, then every variable would necessarily be divisible into the said two factors, is much more difficult, and will be considered in a subsequent paper. Its solution has been achieved by Garnett, and by Spearman.[2]

---

[1] *American Journal of Psychology*, Vol. XV, 1904, pp. 268–272.

[2] Garnett, M.: *Proceedings of Royal Society*, Vol. LXXXXVI, A, 1919.
Spearman, C.: *Proceedings of Royal Society*, Vol. CI, A, 1922, and *Psychology Review*, 1920, pp. 167–168.]Appendix I, 3.]

The psychological significance of the tetrad technique is most clearly portrayed in *The Abilities of Man.* It offers a method of analysing the sub-structure of mind in a rigidly scientific and objective manner, and supplies the firm foundation of a meaningful functional psychology. The objective and functional aspects of validity open up a most promising approach to the entire field of mental tests, and make possible the true cooperation of Individual Differences and General Psychology. This being so, the practical applications of the factor-analysis method are obvious.

(*h*) *The Reliability of the Tetrad Criterion.*—We shall here attempt to analyze the formula given by Spearman for determining the standard deviation of the tetrad difference.[1] Taking, as a typical tetrad difference, $r_{13}r_{24} - r_{23}r_{14}$, the standard deviation has the picturesque value

$$\frac{1}{N}[r^2_{13} + r^2_{14} + r^2_{23} + r^2_{24} - 2(r_{12}r_{13}r_{23} + r_{12}r_{14}r_{24} + r_{13}r_{14}r_{34} + r_{23}r_{24}r_{34}) + 4r_{13}r_{14}r_{23}r_{24}]$$

Let T_{1324} (or $r_{13}r_{24} - r_{23}r_{14}$) be a tetrad representing the entire population, and further, let us assume that the true tetrad difference equals zero. In order to obtain an expression for the reliability of any observed tetrad difference, we must replace each r by $r + dr$, where dr is the ratio of change that r undergoes due to inadequacies of sampling. It may increase r, or it may decrease it.

$$T_{1324} = r_{13}r_{24} - r_{23}r_{14} \qquad (A)$$

$$\begin{aligned} T_{1324} + d_{1324} &= (r_{13} + dr_{13})(r_{24} + dr_{24}) - (r_{23} + dr_{23})(r_{14} + dr_{14}) \\ &= r_{13}r_{24} + dr_{13}r_{24} + dr_{24}r_{13} + dr_{13}dr_{24} - r_{23}r_{14} - \\ &\qquad dr_{14}r_{23} - dr_{23}r_{14} - dr_{23}dr_{14} \end{aligned}$$

Subtracting T_{1324} from both sides, and neglecting the second degree terms of d's, (which are of the order $1/\sqrt{N}$ and small, when N is large) we have

$$d_{1324} = dr_{13}r_{24} + dr_{24}r_{13} - dr_{14}r_{23} - dr_{23}r_{14} \qquad (B)$$

Squaring Equation (*B*),

$$\begin{aligned} d^2_{1324} = {} & \bar{dr}^2_{24}r^2_{13} + \bar{dr}^2_{13}r^2_{24} + \bar{dr}^2_{14}r^2_{23} + \bar{dr}^2_{23}r^2_{14} \qquad (C) \\ & + 2dr_{24}dr_{13}r_{13}r_{24} - 2dr_{24}dr_{14}r_{13}r_{23} \\ & - 2dr_{24}r_{13}dr_{23}r_{14} - 2dr_{13}dr_{14}r_{24}r_{23} \\ & - 2dr_{13}dr_{23}r_{24}r_{14} + 2dr_{14}dr_{23}r_{14}r_{23} \end{aligned}$$

[1] Abilities of Man, Appendix II, 3.

Summing, and dividing by N (the population). Equation (C) becomes

$$\frac{\Sigma d^2_{1324}}{N} = \frac{\Sigma \bar{d}r^2_{24}r^2_{13}}{N} + \frac{\Sigma \bar{d}r^2_{13}r^2_{24}}{N} + \frac{\Sigma \bar{d}r^2_{14}r^2_{23}}{N} \qquad (D)$$
$$+ \frac{\Sigma \bar{d}r^2_{23}r^2_{14}}{N} + 2\frac{\Sigma dr_{24}dr_{13}r_{13}r_{24}}{N}$$
$$- \frac{2\Sigma dr_{24}dr_{14}r_{13}r_{23}}{N} - 2\frac{\Sigma dr_{24}dr_{23}r_{13}r_{14}}{N}$$
$$- \frac{2\Sigma dr_{13}dr_{14}r_{24}r_{23}}{N} - 2\frac{\Sigma dr_{13}dr_{23}r_{14}r_{24}}{N}$$
$$+ 2\frac{\Sigma dr_{14}dr_{23}r_{23}r_{14}}{N}$$

To Simplify

$$\frac{\Sigma \bar{d}r^2_{24}r^2_{13}}{N} = \sigma^2_{r_{24}}r^2_{13} = \frac{(1 - I^2_{24})^2 r^2_{13}}{N}$$

The next three terms may be simplified by a similar process.

To simplify $2\frac{\Sigma dr_{24}dr_{13}r_{13}r_{24}}{N}$, multiply numerator and denominator by $\sigma_{r_{24}}\sigma_{r_{13}}$, and we get

$$2\frac{\Sigma dr_{24}dr_{13}}{N\sigma_{r_{24}}\sigma_{r_{13}}} \cdot r_{13}r_{24}\sigma_{r_{13}}\sigma_{r_{24}} = 2r_{r_{24}r_{13}}r_{13}r_{24}\sigma_{r_{24}}\sigma_{r_{13}}$$

The next five terms may be simplified similarly.

We may now write for Equation (D)

$$\frac{\Sigma d^2_{1324}}{N} = \sigma^2_{r_{24}}r^2_{13} + \sigma^2_{r_{13}}r^2_{24} + \sigma^2_{r_{14}}r^2_{23} + \sigma^2_{r_{23}}r^2_{14} \qquad (E)$$
$$+ 2r_{r_{24}r_{13}}r_{13}r_{24}\sigma_{r_{24}}\sigma_{r_{13}}$$
$$- 2r_{r_{24}r_{14}}r_{13}r_{23}\sigma_{r_{24}}\sigma_{r_{14}}$$
$$- 2r_{r_{24}r_{23}}r_{13}r_{14}\sigma_{r_{24}}\sigma_{r_{23}}$$
$$- 2r_{r_{13}r_{14}}r_{24}r_{23}\sigma_{r_{13}}\sigma_{r_{14}}$$
$$- 2r_{r_{13}r_{23}}r_{24}r_{14}\sigma_{r_{13}}\sigma_{r_{23}}$$
$$+ 2r_{r_{14}r_{23}}r_{23}r_{14}\sigma_{r_{14}}\sigma_{r_{23}}$$

Values for r_{rr} may be obtained by utilizing the formulae of Pearson and Filon. Values for σ_r are easily deduced from the generally accepted reliability formula, $\sigma_r = \frac{1 - r^2}{\sqrt{n}}$.

The Pearson-Filon formulae[1] for correlation between correlations (r_{rr}) are:

$$r_{r_{12}r_{34}} = \frac{\left\{\begin{array}{l}(r_{13} - r_{12}r_{23})(r_{24} - r_{23}r_{34}) + (r_{14} - r_{13}r_{34})(r_{23} - r_{21}r_{13}) \\ +(r_{13} - r_{14}r_{34})(r_{24} - r_{12}r_{14}) + (r_{14} - r_{12}r_{24})(r_{23} - r_{24}r_{34})\end{array}\right.}{2(1 - r^2_{12})(1 - r^2_{34})} \quad (I)$$

$$r_{r_{12}r_{13}} = r_{23} - \frac{r_{12}r_{13}(1 + 2r_{12}r_{23}r_{13} - r^2_{12} - r^2_{23} - r^2_{13}}{2(1 - r^2_{12})(1 - r^2_{13})} \quad (II)$$

We may now write, instead of Equation (*E*):

$$\sigma^2_{\text{Tetrad}} = \frac{(1 - r^2_{24})^2 r^2_{13}}{N} + \frac{(1 - r^2_{13})^2 r^2_{24}}{N} + \frac{(1 - r^2_{14})r^2_{23}}{N} + \frac{(1 - r^2_{23})^2 r^2_{14}}{N} \quad (F)$$

$$+ 2\left\{\frac{\begin{array}{l}(r_{12} - r_{24}r_{14})(r_{34} - r_{14}r_{13}) + (r_{23} - r_{12}r_{13})(r_{14} - r_{24}r_{12}) \\ + (r_{12} - r_{23}r_{13})(r_{34} - r_{24}r_{23}) + (r_{23} - r_{24}r_{34})(r_{14} - r_{13}r_{34})\end{array}}{2(1 - r^2_{24})(1 - r^2_{13})}\right\} r_{13}r_{24}\frac{(1 - r^2_{24})(1 - r^2_{13})}{\sqrt{N}\sqrt{N}}$$

+(Similar expansion of the 6th term in equation *E*)

$$-2\left\{r_{12} - \frac{r_{24}r_{14}(1 + 2r_{24}r_{12}r_{14} - r^2_{24} - r^2_{12} - r^2_{14})}{2(1 - r^2_{24})(1 - r^2_{14})}\right\} r_{13}r_{23} \cdot \frac{(1 - r^2_{24})(1 - r^2_{14})}{\sqrt{N}\sqrt{N}}$$

−(Similar expansion of 3rd, 4th and 5th terms in equation *E*).

By expanding equation (*F*) and assuming that the tetrad differences (true) are zero, so that $r_{13}r_{24} = r_{12}r_{34} = r_{14}r_{23}$, we may then collect terms, and, by ordinary arithmetic obtain the following:

$$\sigma^2_{\text{Tetrad}} = \frac{1}{N}[r^2_{13} + r^2_{14} + r^2_{23} + r^2_{24} - 2(r_{13}r_{12}r_{23} + r_{12}r_{14}r_{24} + r_{13}r_{14}r_{34} + r_{23}r_{24}r_{34}) + 4r_{13}r_{14}r_{23}r_{24}] \quad (G)$$

(*i*) *Freedom of the Tetrad Criterion from the Influence of Attenuation* (*Abilities, I*, 4). Granting the fact that the tetrad criterion holds for "corrected" correlations, does it also hold for the corresponding "raw" Correlations?

---

[1] *Philosophical Transactions*, Vol. CXCI A, 1898, pp. 229–311.

Let x_1, y_1, z_1, w_1 be any four uncorrected correlations forming a rectangle in a table of correlations; *e.g.*

	X	*Y*	*Z*	*W*	*E*
X	–				
Y		–			
Z	$Y_{x_1z_1}$	$Y_{y_1z_1}$	–		
W	$Y_{x_1w_1}$	$Y_{y_1w_1}$		–	
E					–

Let x, y, z, w denote the corresponding corrected or true values.

We have assumed that $r_{xz} \cdot r_{yw} = r_{yz} \cdot r_{xw}$

Our question is, will the equation hold for the raw values, *i.e.* does

$$r_{x_1z_1} \cdot r_{y_1w_1} = r_{y_1z_1} \cdot r_{x_1w_1}?$$

Let us first reduce the terms of each expression to deviation form, and then compare the results:

$$r_{xz} = \frac{\Sigma xz}{\sqrt{\Sigma x^2 \Sigma z^2}}, \qquad r_{x_1z_1} = \frac{\Sigma x_1 z_1}{\sqrt{\Sigma x_1{}^2 \cdot \Sigma z_1{}^2}}$$

and similarly for the other terms.

For true values,

$$r_{xz} \cdot r_{yw} = r_{yz} \cdot r_{xw} \tag{1}$$

$$\therefore \frac{\Sigma xz \cdot \Sigma yw}{\sqrt{\Sigma x^2 \Sigma z^2 \Sigma y^2 \Sigma w^2}} = \frac{\Sigma yz \cdot \Sigma xw}{\sqrt{\Sigma y^2 \Sigma z^2 \Sigma x^2 \Sigma w^2}} \tag{2}$$

$$\therefore \Sigma xz \cdot \Sigma yw = \Sigma yz \cdot \Sigma xw \tag{3}$$

In the analysis of the correction formula, (Sub-section *e*) it was shown that the sum of the products of the pairs of deviations of the true values of two variables, equals the sum of the products of the pairs of deviations of the observed values (p. 204).

Therefore:

$$\Sigma xz = \Sigma x_1 z_1 \qquad \Sigma xw = \Sigma x_1 w_1$$

$$\Sigma yw = \Sigma y_1 w_1 \qquad \Sigma yz = \Sigma y_1 z_1$$

Substituting equivalent observed values for true values in equation 3, we get:

$$\Sigma x_1 z_1 \cdot \Sigma y_1 w_1 = \Sigma y_1 z_1 \cdot \Sigma x_1 w_1$$

Dividing through by $\sqrt{\Sigma x_1^2 \Sigma y_1^2 \Sigma z_1^2 \Sigma w_1^2}$

$$\frac{\Sigma x_1 z_1 \cdot \Sigma y_1 w_1}{\sqrt{\Sigma x_1^2 \Sigma z_1^2 \Sigma y_1^2 \Sigma w_1^2}} = \frac{\Sigma y_1 z_1 \cdot \Sigma x_1 w_1}{\sqrt{\Sigma y_1^2 \Sigma z_1^2 \Sigma x_1^2 \Sigma w_1^2}}$$

or

$$r_{x_1 z_1} \cdot r_{y_1 w_1} = r_{y_1 z_1} \cdot r_{x_1 w_1}$$

In other words, the criterion holds for uncorrected correlations whenever it does so for the corresponding true values. (In Section B, (*a*), an alternative method of proof is offered, the problem being approached there from the opposite angle, namely, 'given that the tetrad equation is satisfied with raw values, does it follow that it holds for the true correlations corresponding?)

SECTION B

(*a*) *Alternative Method of Correcting for Attenuation.*[1]—Method.—Several measurements are taken of each individual in the two abilities to be correlated. The measurements for each individual, and for each ability, are then arbitrarily divided into two groups, so that the difference between the group averages for an individual is purely "accidental." Such a division may be made by taking measurements 1, 3, 5, etc. for group one, 2, 4, 6, etc. for group two.

We many safely assume that the "accidental" errors in one ability are not correlated with those in the other, and that such errors are not correlated with the true underlying values corresponding to the measurements observed. We may also assume that the averages of the true underlying values of the two groups and of the sum of the two groups are all equal. Thirdly we may assume that the sum of the squares of the individual averages in one group is not very different in magnitude from the sum of the squares of the individual averages in group two. Symbols employed in this analysis (deviation form)

Let x denote the average value of the underlying the true measurements.

x_1 the average of the measures for an individual in group 1

x_1^3 the regular underlying measure corresponding to x_1

d_1 the accidental error

x_{12} the average of the obtained measures in both groups.

x^1_{12} the average of underlying true measures for this individual in both groups.

d_{12} the accidental error.

[1] Spearman, C.: *British Journal of Psychology*, Vol. III, p. 288.

Let $x_2x_2{}^1d_2$ denote similar terms for the same individual in group 2.

Let $y_1y_1{}^1e_1y_{12}y_{12}{}^1e_{12}y_2y_2{}^1e_2$ denote similar terms for the same individual in the two groups; obtained from his measurements in respect of the other ability (y variable).

We may now state our assumptions and equations concisely:

Assumptions:

(1) $x^1{}_1 = x^1{}_2 = x^1{}_{12}$ (see page 274 in the article cited)

(2) $x = x^1{}_{12}$ since the average of the various deviations from the average (x^1) must equal zero.

(3) Σxe, Σyd, Σde, etc.

(4) $\Sigma d_1{}^2$ and $\Sigma d_2{}^2$ are very nearly equal.
and $\Sigma x_1{}^2$ and $\Sigma x_2{}^2$ are very nearly equal.

Equations given:

$$x_1 = x^1{}_1 + d_1 = x^1{}_{12} + d_1 \text{ and since } x^1{}_{12} = x$$
$$x_1 = x + d_1$$
$$x_{12} = x + d_{12}$$
$$x_2 = x + d_2$$
$$y_1 = y + e_1$$
$$y_2 = y + e_2$$
$$y_{12} = y + e_{12}$$

Then

$$x_{12}y_{12} = (x + d_{12})(y + e_{12})$$
$$\Sigma x_{12}y_{12} = \Sigma(x + d_{12})(y + e_{12}) \qquad \text{(i)}$$
$$x_1x_2 = (x + d_1)(x + d_2)$$
$$\Sigma x_1x_2 = \Sigma(x + d_1)(x + d_2) \qquad \text{(ii)}$$
$$y_1y_2 = (y + e_1)(y + e_2)$$
$$\Sigma y_1y_2 = \Sigma(y + e_1)(y + e_2) \qquad \text{(iii)}$$

Multiplying (ii) by (iii) and taking the square root:

$$[\Sigma x_1x_2 \cdot \Sigma y_1y_2]^{1/2} = [\Sigma(x + d_1)(x + d_2)\Sigma(y + e_1)(y + e_2)]^{1/2} \qquad \text{(iv)}$$

Dividing (i) by (iv)

$$\frac{\Sigma x_{12}y_{12}}{[\Sigma x_1x_2\Sigma y_1y_2]^{1/2}} = \frac{\Sigma(x + d_{12})(y + e_{12})}{[\Sigma(x + d_1)(x + d_2)\Sigma(y + e_1)(y + e_2)]^{1/2}}$$
$$= \frac{\Sigma(xy + xe_{12} + yd_{12} + e_{12}d_{12})}{[\Sigma(x^2 + xd_2 + xd_1 + d_1d_2)\Sigma(y^2 + ye_1 + ye_2 + e_1e_2)]^{1/2}}$$
$$= \text{(By assumption 3)}\frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = r_{xy} \qquad \text{(v)}$$

We now have an expression for r_{xy} in terms of group deviations. Our task is to reduce the left hand side of equation (v) to a convenient form.

If we can express r_{xy} as a function of $r_{x_{12}y_{12}}$, we shall have a formula of practical value which will correct for accidental errors.

It is clear that the denominator of equation (v) (left side) must be reduced to the form $[\Sigma x^2_{12}\Sigma y^2_{12}]^{\frac{1}{2}}$ if we are to get r_{xy} in terms of $r_{x_{12}y_{12}}$.

$$\frac{\Sigma x_{12}y_{12}}{[\Sigma x_1x_2\Sigma y_1y_2]^{\frac{1}{2}}} = \frac{\Sigma x_{12}y_{12}}{[\Sigma x^2_{12}\Sigma y^2_{12}]^{\frac{1}{2}}} \cdot \frac{[\Sigma x^2_{12}\Sigma y^2_{12}]^{\frac{1}{2}}}{[\Sigma x_1x_2\Sigma y_1y_2]^{\frac{1}{2}}}$$

$$\therefore r_{xy} = r_{x_{12}y_{12}}\left(\frac{\Sigma x^2_{12}\Sigma y^2_{12}}{\Sigma x_1x_2\Sigma y_1y_2}\right) = r_{x_{12}y_{12}}\sqrt{\frac{\Sigma x^2_{12}}{\Sigma x_1x_2}}\sqrt{\frac{\Sigma y^2_{12}}{\Sigma y_1y_2}}$$

But

$$\sqrt{\frac{\Sigma x^2_{12}}{\Sigma x_1x_2}} = \sqrt{\frac{\Sigma\left(\frac{x_1+x_2}{2}\right)^2}{\Sigma x_1x_2}}$$

$$= \sqrt{\frac{\Sigma x_1{}^2 + 2\Sigma x_1x_2 + \Sigma x_2{}^2}{4\Sigma x_1x_2}}$$

$$= \sqrt{\frac{1}{2}\left(1 + \frac{\Sigma x_1{}^2 + \Sigma x_2{}^2}{2\Sigma x_1x_2}\right)}$$

By assumption 4, $\Sigma x_1{}^2 = \Sigma x_2{}^2$.

$$\therefore \Sigma x_1{}^2 + \Sigma x_2{}^2 = 2\Sigma x_1{}^2 = 2\sqrt{\Sigma x_1\Sigma x_1} = 2\sqrt{\Sigma x_1\Sigma x_2}$$

$$\therefore \sqrt{\frac{\Sigma x^2_{12}}{\Sigma x_1x_2}} = \sqrt{\frac{1}{2}\left(1 + \frac{2\sqrt{\Sigma x_1{}^2\Sigma x_2{}^2}}{2\Sigma x_1x_2}\right)}$$

$$= \sqrt{\frac{1}{2}\left(1 + \frac{1}{\frac{\Sigma x_1x_2}{\sqrt{\Sigma x_1{}^2\Sigma x_2{}^2}}}\right)}$$

$$= \sqrt{\frac{1}{2}\left(1 + \frac{1}{r_{x_1x_2}}\right)}$$

Similarly

$$\sqrt{\frac{\Sigma y^2_{12}}{\Sigma y_1y_2}} = \sqrt{\frac{1}{2}\left(1 + \frac{1}{r_{y_1y_2}}\right)}$$

$$\therefore r_{xy} = r_{x_{12}y_{12}}\sqrt{\frac{1}{2}\left(1 + \frac{1}{r_{x_1x_2}}\right)}\sqrt{\frac{1}{2}\left(1 + \frac{1}{r_{y_1y_2}}\right)}$$

$$= r_{x_{12}y_{12}} \cdot \frac{1}{2} \cdot \sqrt{\left(1 + \frac{1}{r_{x_1x_2}}\right)\left(1 + \frac{1}{r_{y_1y_2}}\right)}$$ (given thus in the article cited)

$$= r_{x_{12}y_{12}}\frac{(1 + r_{x_1x_2})^{\frac{1}{2}}(1 + r_{y_1y_2})^{\frac{1}{2}}}{2r^{\frac{1}{2}}_{x_1x_2}r^{\frac{1}{2}}_{y_1y_2}}$$ (given thus in '*Abilities*' I. 4.)

It is again obvious that, if each observed correlation is to be multiplied by a factor as indicated in this last equation, the Tetrad Difference obtained with observed correlations will not be affected by the correction.

(*b*) *Alternative Method of Deriving the Tetrad Criterion: Partial Correlation.*—Before showing how the Tetrad Equation is related to Yule's well-known formula for Partial Correlations, it is advisable to indicate how that formula may be derived.

In the case of two variables, there are two regression equations:

$x = r_{xy}\frac{\sigma_x}{\sigma_y}y$—the relation of any mean x to its particular y.

$y = r_{yx}\frac{\sigma_y}{\sigma_x}x$—the relation of any mean y to its particular x.

$r_{xy}\frac{\sigma_x}{\sigma_y}$ is the regression coefficient of x on y.

$r_{yx}\frac{\sigma_y}{\sigma_x}$ is the regression coefficient of y on x.

r^2_{xy} is the product of the regression coefficients.

Let us suppose that there is a third variable (z) which influences or is correlated with x and y—and is active. Our problem then is to eliminate the influence of z. We must therefore first take account of its influence on the degree of correlation between x and y.

Let us, accordingly, develop:

(1) an expression for the regression coefficient of x on y independent of z, but in terms of known values, *i.e.* in terms of the two-variable coefficient r_{xy}.

(2) an expression for the regression coefficient of y on x similarly.

(3) by the product of these two coefficients, an expression for the square of the correlation between x and y with z controlled, ($r^2_{xy\cdot z}$) in terms of r_{xy}.

In the case of two variables, the means of the rows of x's were represented by the "best fitting" line. With three variables, these same means, which were assembled in a line, are now scattered up and down in a plane passing through that line. Their various positions in that plane mark their z-ness. Approaching the problem by the same method as before, we see that, for a single x deviation in a row

$$x = r^1_{xy\cdot z}\frac{\sigma_x}{\sigma_y}y + r^1_{xz\cdot y}\frac{\sigma_x}{\sigma_z}z.$$

For a row

$$\frac{\Sigma x}{n} = \frac{\Sigma r^1{}_{xy\cdot z}}{n}\frac{\sigma_x}{\sigma_y}y + \frac{\Sigma r^1{}_{xz\cdot y}}{n}\frac{\sigma_x}{\sigma_z}z.$$

$$\text{Mean } x = r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}}y + r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}z. \tag{1}$$

Here we have an expression for the sum total of the mean x deviation with respect to both y and z. The term $r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}}$ expresses the relation between the x deviation of a mean and its particular y deviation for all values of z, *i.e.* it is independent of z. Shifting up and down the plane does not alter the x or y deviation. It simply adds another deviation, that of z. Likewise $r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}$ expresses the relation between a mean x deviation and its particular z deviation independent of y.

In equation (1) we have an expression for the regression of x on y and z stated in two terms—independent of z in the one term, independent of y in the other. We need an equivalent expression for the regression of x on y and z, but in terms of r_{xy} and r_{xz}, with z and y exerting an influence. These are correlations which we can readily obtain experimentally.

Let us then reduce $r_{xy\cdot z}\frac{\sigma_{x.yz}}{\sigma_{y\cdot xz}}$ and $r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z.xy}}$ to $r_{xy}\frac{\sigma_x}{\sigma_y}$ and $r_{xz}\frac{\sigma_x}{\sigma_y}$; *i.e.* to known terms.

Transposing, in equation (1)

$$r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}}y = x - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}z \tag{2}$$

$$r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}}y^2 = xy - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}yz \tag{3}$$

$$r_{y\cdot xz}\frac{\sigma_{x\cdot yz}}{\sigma_{x\cdot yz}}\Sigma y^2 = \Sigma xy - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}}\Sigma yz \tag{4}$$

But

$$\Sigma_y{}^2 = n\sigma_y{}^2.$$

$$\therefore r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = \frac{\Sigma_{xy}}{n\sigma_y{}^2} - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}\frac{\Sigma_{xz}}{n\sigma_y{}^2} \tag{5}$$

$$r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = \frac{\Sigma_{xy}}{n\sigma_x\sigma_y}\cdot\frac{\sigma_z}{\sigma_y} - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}\frac{\Sigma_{yz}}{n\sigma_y\sigma_z}\frac{\sigma_z}{\sigma_y} \tag{6}$$

$$r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = r_{xy}\frac{\sigma_z}{\sigma_y} - r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}r_{yz}\frac{\sigma_z}{\sigma_y} \tag{7}$$

But

$$r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot yz}} = r_{xz}\frac{\sigma_x}{\sigma_z} - r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{x\cdot yz}}r_{yz}\frac{\sigma_y}{\sigma_z}$$

(deduced from equation (7) by an obvious shifting of subscripts.)

Substituting accordingly for $r_{xz\cdot y}\frac{\sigma_{x\cdot yz}}{\sigma_{z\cdot xy}}$ in equation (7)

$$r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = r_{xy}\frac{\sigma_x}{\sigma_y} - r_{yz}\frac{\sigma_z}{\sigma_y}\cdot r_{xz}\frac{\sigma_x}{\sigma_z} + r^2{}_{yz}\frac{\sigma_z}{\sigma_y}\frac{\sigma_y}{\sigma_z}r_{xy\cdot z}\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} \tag{8}$$

Transposing the unknown "partial" to the left:

$$(r_{xy\cdot z} - r_{yz}r_{xy\cdot z})\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = (r_{xy} - r_{xz}r_{yz})\frac{\sigma_x}{\sigma_y} \tag{9}$$

or

$$y_{xy\cdot z}(1 - r^2{}_{yz})\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = (r_{xy} - r_{xz}r_{yz})\frac{\sigma_x}{\sigma_y}$$

$$\therefore r_{xy\cdot z}\cdot\frac{\sigma_{x\cdot yz}}{\sigma_{y\cdot xz}} = \frac{r_{xy} - r_{xz}\cdot r_{yz}}{1 - r^2{}_{yz}}\frac{\sigma_x}{\sigma_y} \tag{10}$$

The Regression Coefficient of x on y, independent of z, and in terms of known correlations.

By obvious shifting of subscripts, the regression coefficient of y on x is

$$r_{yz\cdot x}\frac{\sigma_{y\cdot xz}}{\sigma_{x\cdot yz}} = \frac{r_{yz} - r_{xz}r_{yz}}{1 - r^2{}_{xz}}\frac{\sigma_y}{\sigma_x} \tag{11}$$

$$\therefore r^2{}_{xy\cdot z} = \frac{(r_{xy} - r_{xz}r_{yz})^2}{(1 - r^2{}_{xz})(1 - r^2{}_{yz})}$$

$$\therefore r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1 - r^2{}_{xz}}\sqrt{1 - r^2{}_{yz}}}.$$

The partial regression equations may now be written in terms of the regression coefficients just obitained:

$$x = \frac{r_{xy} - r_{xz}r_{yz}}{1 - r^2{}_{yz}}\frac{\sigma_x}{\sigma_y}\cdot y + \frac{r_{xz} - r_{xy}r_{xz}}{1 - r^2{}_{yz}}\frac{\sigma_y}{\sigma_z}\cdot z \tag{1}$$

$$y = \frac{r_{yz} - r_{yz}r_{xz}}{1 - r^2{}_{xz}}\frac{\sigma_y}{\sigma_z}\cdot x + \frac{r_{yz} - r_{yx}r_{xz}}{1 - r^2{}_{xz}}\frac{\sigma_y}{\sigma_z}\cdot z \tag{2}$$

$$z = \frac{r_{xz} - r_{xy}r_{zy}}{1 - y^2{}_{yz}}\frac{\sigma_z}{\sigma_x}\cdot x + \frac{r_{zy} - r_{yx}r_{xz}}{1 - r^2{}_{xz}}\frac{\sigma_z}{\sigma_x}\cdot y \tag{3}$$

These regression equations are useful for predicting the values in one variable from given values in the other two.

(The correlation between this predicted value and the observed value is the Multiple Coefficient of Correlation.)

Let us now show the validity of the Tetrad Equation by applying the formula derived above.

Let $r_{x_1y_1 \cdot g}$ denote the correlation that x_1 would have with y_1 if the influence of g were eliminated. Then,

$$r_{x_1y_1 \cdot g} = \frac{r_{x_1y_1} - r_{x_1g}r_{y_1g}}{\sqrt{1 - r^2_{x_1g}}\sqrt{1 - r^2_{y_1g}}}$$

But

$$r_{x_1y_1 \cdot g} = 0$$

$$\therefore r_{x_1 \cdot y_1} = r_{x_1g}r_{y_1g} \qquad (1)$$

Similarly

$$r_{x_2y_1} = r_{x_2g}r_{y_1g} \qquad (2)$$

$$\therefore \frac{r_{x_1y_1}}{r_{x_2y_1}} = \frac{r_{x_1g}}{r_{x_2g}} \qquad (A)$$

Similarly it may be shown that

$$\frac{r_{x_2y_2}}{r_{x_1y_2}} = \frac{r_{x_2g}}{r_{x_1g}} \qquad (B)$$

From (A) and (B) it follows that

$$\frac{r_{x_1y_1}}{r_{x_2y_1}} = \frac{r_{x_1y_2}}{r_{x_2y_2}},$$

or

$$r_{x_1y_1}r_{x_2y_2} = r_{x_2y_1}r_{x_1y_2}. \quad \text{(Appendix I. 2.)}$$

SECTION C SYMBOLS USED IN THIS PAPER

Except where otherwise indicated, the following interpretation of symbols holds good:

- X a true score, variable 1.
- X_m the mean true score, variable 1.
- x the deviation of a true score $(X - X_m)$, variable 1.
- Y a true score, variable 2.
- Y_m the mean true score, variable 2.
- y the deviation of a true score $(Y - Y_m)$, variable 2.
- r the coefficient of correlation between true values of variables 1 and 2.

X_1 an observed score, variable 1.
x_1 the deviation of an observed score, $(X - X_{1m})$ variable 1.
Y_1, y_1 similarly for variable 2.
r_1 the coefficient of correlation between observed values of variables 1 and 2.
e_{x_1} the error in x_1.
e_{y_1} the error in y_1.

BIBLIOGRAPHY FOR TECHNICAL STUDY OF THE TWO-FACTOR THEORY

Spearman, C.: The Measurement of Association between Two Things. *American Journal of Psychology*, Vol. XV, 1904, pp. 72–101.

——: General Intelligence,—Objectively Measured and Determined. *American Journal of Psychology*, Vol. XV, 1904, pp. 201–293.

Krueger, F. and Spearman, C.: Die Korrelation zwischen verschiedenen geistigen Leistungsfahigkeiten. *Zeitschrift für Psychologig*, Vol. XLIV, 1906, pp. 50–114.

Hart, B. and Spearman, C.: General Ability, Its Existence and Nature. *British Journal of Psychology*, Vol. V, 1912, pp. 51–84.

Hart, B. and Spearman, C.: Mental Tests of Dementia. *Journal of Abnormal Psychology*, Vol. IX, 1914, pp. 217–264.

Spearman, C.: The Theory of Two Factors. *Psychological Reviews*, Vol. XXI, 1914, pp. 101–115.

——: Heredity of Abilities. *Eugenic Reviews*, Vol. VI, 1914, pp. 219–237.

Garnett, J. C. M.: General Ability, Cleverness and Purpose. *British Journal of Psychology*, Vol. IX, 1919, pp. 345–366.

——: The Single General Factor in Dissimilar Mental Measurements. *British Journal of Psychology*, Vol. X, 1920, pp. 242–258.

Spearman, C.: Manifold Theories of "The Two Factors." *Psychological Reviews*, Vol. XXVII, 1920, pp. 159–172.

——: Correlation between Arrays in a Table of Correlations. *Proceedings of Royal Society of London*, 1922, 101A, pp. 94–100.

——: Recent Contributions to the Theory of Two Factors. *British Journal of Psychology*, Vol. XXIII, 1922, pp. 26–30.

——: The Nature of "Intelligence" etc. Macmillan, London, 1923.

—— and Holzinger, K.: The Sampling Error in the Theory of Two Factors. *British Journal of Psychology*, Vol. XV, 1924, pp. 17–19.

—— and Holzinger, K.: Note on the Sampling Error of the Tetrad Differences. *British Journal of Psychology*, Vol. XVI, 1925, pp. 86–88.

——: Some Issues in the Theory of g. *Proceedings of British Association of Advanced Science*, 1925, Sect. J. pp. 174–181.

——: Material versus Abstract Factors in Correlation. *British Journal of Psychology*, Vol. XVII, 1927, pp. 322–326.

——: "The Abilities of Man." Macmillan, London, 1927.

Dodd, Stuart C.: The Theory of Factors. *Psychological Reviews*, Vol. XXXV, 1928, pp. 211–234 and 261–279.

Spearman, C.: Elimination of Accidental Errors. *British Journal of Psychology*, Vol. 3, 1910, pp. 276.

——: The Correlation for Attenuation. *American Journal of Psychology*, Vol. XVIII.

Garnett, J. C. M.: *Proceedings of Royal Society*, Vol. LXXXXVI, A, 1919.

Brown, W. and Thompson, G. H.: "Essentials of Mental Measurement." Cantab. U. P., 1921.

Kelley, T. L.: "Crossroads." Stanford Press, Stanford University, California, 1929.

Spearman, C.: Reply to Kelley. *Journal of Educational Psychology*, 1930.

—— and Holzinger, K.: Probable Error of Tetrad Differences. *British Journal of Psychology*, Vol. XX, 1930, pp. 368–370.

Thompson, Godfrey, H.: A Hierarchy without a General Factor. *British Journal of Psychology*, Vol. VIII, 1916, p. 271, (followed by Spearman's comments).

——: The Proof or Disproof of the Existence of General Ability. *British Journal of Psychology*, Vol. IX, 1919, pp. 321–336.

——: The Hierarchy of Abilities. *Ibid.* Vol. IX, 1919, pp. 337–344.

—— and Garnett, J. C. M.: Joint Note on the above article, *ibid.*, p. 345.

——: On the Cause of Hierarchical Order among the Correlation Coefficients of a Number of Variates Taken in Pairs. *Proceedings of Royal Society of London*, 1919, Vol. XCV A, pp. 400–408.

——: On the Degree of Perfection of Hierarchical Order among Correlational Coefficients. *Biometrica*, Vol. XII, 1919, p. 355, and Vol. XV, 1923, p. 150.

——: The General Factor Fallacy in Psychology. *British Journal of Psychology*, Vol. X, 1920, p. 319.

——: General versus Group Factors in Mental Activities. *Psychological Reviews*, Vol. XXVII, 1920, p. 173.

——: The Tetrad-difference Criterion. *British Journal of Psychology*, Vol. XVII, 1927, pp. 235–255.

——: A Worked-out Example, etc. *British Journal of Psychology*, Vol. XVIII, 1927, pp. 68–76.

——: On the Formation of Structure Diagrams between Four Correlated Variables. *Journal of Educational Psychology*, Vol. XVIII, 1927, pp. 145–158.

Garnett, J. C. M.: On Certain Independent Factors in Mental Measurements. *Proceedings of Royal Society of London*, Vol. XCVI A, 1919, pp. 91–111.

——: The Single General Factor in Dissimilar Mental Measurements. *British Journal of Psychology*, Vol. X, 1920, pp. 242–258.

Garnett, J. C. M.: "Review of 'The Essentials of Mental Measurement', Brain." Vol. XLIV, 1921, pp. 332–337.

Spearman, C.: Further Note on the Theory of Two Factors. *British Journal of Psychology*, Vol. XIII, 1923, pp. 266–270.

——: Agreement on Co-operation. *Journal of Educational Psychology*, Vol. XVI, 1925, pp. 423–425.

——: Factor versus Sampling Theory. (Article to be published shortly.)

THE HEINIS PERSONAL CONSTANT AS A SUBSTITUTE FOR THE IQ

PSYCHE CATTELL

Harvard University

There is a growing dissatisfaction with the IQ among psychologists. As data accumulates the belief that there is a constant tendency for the IQ of the dull child to decrease with increase in age while that of the bright child increases is becoming more prevalent. Dr. Kuhlmann has recommended Dr. Heinis' Personal Constant[1,2] in place of the IQ for use with his tests. He writes:

> That it has been known for some time that the IQ as found by tests that give correct mental ages at all mental levels does not remain constant throughout successive years except for children with initial IQ of 100. Intelligence Quotients below 100 tend to decrease, and above 100 they tend to increase as the child grows older. The amount of change per year depends on the initial IQ and on the chronological age. The total rate during the mental growth period may exceed 25 IQ points. The change is due to the fact that the rate of normal or average mental development slows up with increasing years.[3]

Combining his own data with those obtained by two other investigators, H. Heinis secured data on the mental development of children ranging in age from two to eighteen years. From the material he constructed a formula[4] for a curve of mental growth and a coefficient which he terms a Personal Constant, to be used in place of the IQ. Since the Heinis Personal Constant (H.C.) is based on the same intelligence tests as the IQ, in fact on the mental age derived from the identical test performance, it is, of course, not expected to correct the variable fluctuations in the IQ. It is, however, claimed that it remains more constant than the IQ at the upper and lower levels of intelligence.

In order to make a comparison of the H.C. (Heinis Personal Constant) and the IQ, the records of three groups of children were

---

[1] Heinis: H. La loi du Developpement Mental. *Archives de Psychol.*, No. 74, 1924.

[2] Heinis H.: A Personal Constant. *Journal of Educational Psychology*, March, 1928.

[3] "Instruction Manual for the Kuhlmann-Anderson Intelligence Tests." (Forthcoming revision.)

[4] $y = 429\left(1 - e^{-\frac{X}{6.675}}\right)$.

taken from the files of the Harvard Growth Study.[1] The first group contained those cases whose IQ's during the eight years averaged 120 or higher; the second group those with IQ's averaging between 95 and 104, inclusive; and the third, dull children with average IQ's below 80. The average number of tests for each child was 18.2 or 2.4 for each year. The tests used were:

First year:
- Dearborn General Examination A
- Stanford-Binet Individual Examination (to about one-third the cases)

Second year:
- Dearborn General Examination A
- Stanford-Binet Individual Examination (to about two-thirds the cases)

Third year:
- Dearborn General Examination A
- Otis Primary Intelligence Test, Form A

Fourth year:
- Dearborn General Examination C
- Otis Primary Intelligence Test, Form A

Fifth year:
- Dearborn General Examination C
- Otis Self-administering Test of Mental Ability, Form A

Sixth year:
- Dearborn General Examination C
- Otis Self-administering Test of Mental Ability, Form B

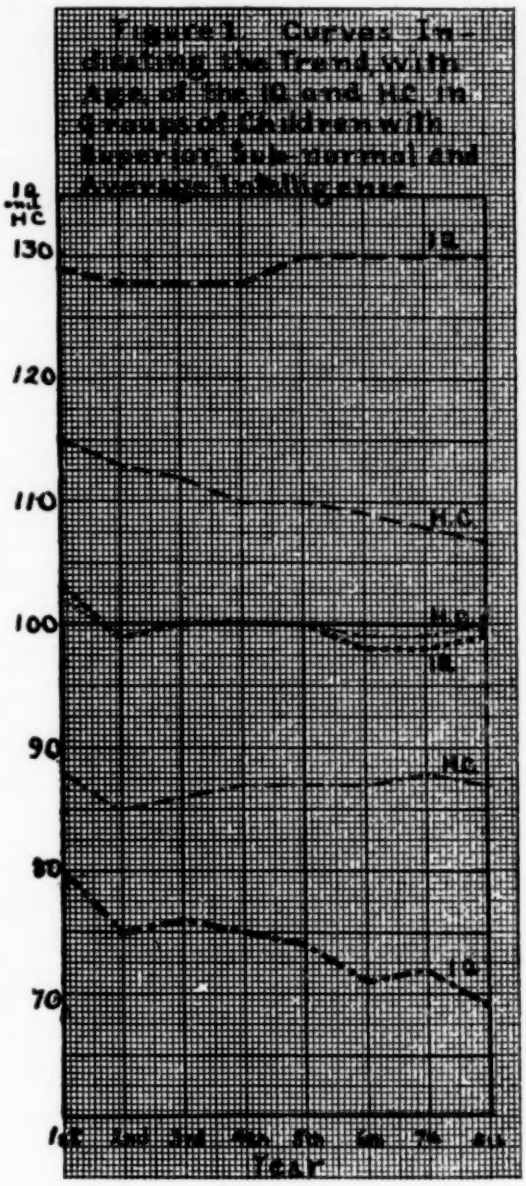
Seventh year:
- Dearborn General Examination, Old Form 4
- Dearborn General Examination, Old Form 5
- Haggerty Intelligence Tests, Delta 2
- Stanford-Binet Individual Examination

Eighth year:
- Haggerty Intelligence Examination, Delta 2
- Terman Group Test of Mental Ability, Form A or B

In order to avoid, as far as possible, variations in the test results due to differences in standardization and the effects of practice each

[1] The Growth Study was started at Harvard University under the direction of Walter F. Dearborn in 1922–1923. During the first year over 3500 children were given physical, intelligence and educational examinations, in so far as they can be located these same children have been examined every year since that date.

group test IQ was equated to the Stanford-Binet.[1] The equated IQ's for each child were then averaged. In about one half of the cases there was only one IQ available for each of the first and second years.



Except in a few scattered instances every child had at least two tests during each of the remaining six years.

Thirty-seven cases with IQ's of one hundred twenty or over, all of North European descent, were available. The files contained

[1] The tables used in equating the tests were loaned to the writer by Professor John Ratcliffe of Tufts College who is preparing, for publication, a set of percentile tables for equating IQ's based on some 2000 cases.

approximately the same number of records of children with IQ's below eighty as of children with IQ's of one hundred twenty and above, but the attendance in school of the dull pupils is less regular and only eighteen of North European descent had had an intelligence test for each of the eight years. In order to increase the size of the group twenty-two Italians and two Jews were added, forming a group of forty-two cases. The average IQ and the general trend of the IQ curves of the pupils of Italian and of North European descent were almost identical. The group of pupils with average intelligence were all of North European descent and contained twenty cases. The mental age for each child for each year was determined from the C.A. and the average of the equated IQ's; the Heinis Constant from the growth units given in the revised manual for the Kuhlmann-Anderson Intelligence Tests.[1]

Table I and the upper heavy broken line in Fig. 1 indicates for each year the average IQ of the thirty-seven superior children and the light broken line the H.C.'s for the same children. In comparing the constancy of the IQ and the H.C. it must be remembered that the range of the H.C. is less than that of the IQ. The average IQ at the end of the seventh year was only one point or three per cent further above one hundred than it was seven years earlier, on the other hand, the Heinis Constant showed a steady downward trend until at the end of the seventh year it was only fifty per cent as far above one hundred as it was at the beginning of the experiment.

TABLE I.—AVERAGE CHANGE IN INTELLIGENCE QUOTIENTS AND HEINIS CONSTANTS WITH AGE OF DULL, AVERAGE AND BRIGHT CHILDREN

Year	Dull		Average		Bright	
	IQ	H.C.	IQ	H.C.	IQ	H.C.
1	80	88	103	102	129	115
2	75	85	99	99	128	113
3	76	86	100	100	128	112
4	75	87	100	100	128	110
5	74	87	100	100	130	110
6	71	87	98	99	130	109
7	72	88	98	99	130	108
8	69	87	99	100	130	107
Mean.....	74	87	100	100	129	110

[1] At the time of writing the revised manual had not come off the press. Advanced sheets kindly supplied by the publisher were used.

The first question that arises is why do the IQ's of these superior pupils remain on the average constant, when most other studies have indicated an increase with age.[1,2,3] The most probable explanation appears to lie in the fact that the former studies of the constancy of the IQ of bright children which have not been complicated by different tests or the effects of practice have been based on Stanford-Binet IQ. The present study is based mainly on group test IQ's. Stanford-Binet IQ's were included in the averages when they were available, but after the second year the number of Binet IQ's were small except in the seventh year when they were averaged in with three group test IQ's.

Twenty-four of the thirty-seven superior pupils had a Stanford-Binet IQ available for the second year and for either the sixth or seventh year of the study. When this test alone is considered the expected gain in IQ is found. The mean Stanford-Binet IQ of twenty-four cases was one hundred twenty-three when the test was given in the second year and one hundred thirty-one when repeated four or five years later, a gain of eight IQ points. The Heinis Constant for the same pupils based on the Stanford-Binet test changed from one hundred eleven to one hundred nine, a change of two H.C. Points, or the equivalent of a loss of a little less than six IQ points. The fact that the group test IQ's remain constant while the Stanford-Binet IQ rises indicates the possibility that the gain in Stanford-Binet IQ of children of superior intelligence reported in the literature may be a function of the particular test rather than of the IQ technique.

Table I and the dot-dash lines in Fig. 1 gives the average IQ and H.C. for a group of forty-two dull children. The downward trend of the IQ's reported in the literature for the Stanford-Binet test is also evident in these group test IQ's. The curve for the Heinis Constant is almost horizontal, thus meeting, at this level, the claims made for it. But it may be that the downward trend of the IQ's of the dull pupils is due to the fact that the group tests as well as the Binet are increasingly scholastic with advancing years and that the dull group lose out academically. Thus the decrease in IQ may be a real condition, not an artifact and should not be "corrected." The IQ of the dull group

---

[1] Riley, Gordon: "A Comparison of the Personal Constant and IQ." Psychological Clinic, February, 1930.

[2] Cattell, Psyche: Constant changes in the Stanford-Binet IQ. *Journal of Educational Psychology*, October, 1931.

[3] Cattell, Psyche: "Do the Stanford-Binet IQ's of Superior Boys and Girls Tend to Decrease or Increase with Age." (Ready for publication.)

might show constancy if some other avenues of expression were tested (the above explanation was suggested by Dr. W. F. Dearborn).

The center section of Table I and the dotted lines in Fig. 1 give a comparison of the H.C. and IQ for the group of children of average intelligence. Both measures remain consistently near the one hundred mark during the seven-year period. The slightly greater closeness of the H.C. to one hundred is of no significance, it is due to the narrower range of the H.C.

Gordon Riley[1] made a study of the comparative constancy of the Heinis Personal Constant and the Stanford-Binet IQ in three groups of feeble-minded children and one group of superior children. His results are in agreement with the present study in so far as the feeble-minded children are concerned, namely that he finds the IQ to decrease while the H.C. remains constant. His results obtained from a group of thirty children with IQ's averaging one hundred eighteen are in contrast to the above reported group test results of children with IQ's averaging one hundred twenty-nine, in that he found the IQ to increase and the H.C. to remain constant. However, the differences between his results and our Stanford-Binet results are small and it may be that they could be accounted for by the fact that his group of superior children were of the lower order of intelligence than those included in the present study.

As a check of the above emperical findings hypothetical cases were selected with H.C.'s of 65, 70 . . . 135, and the IQ's that would of necessity result if the H.C. were constant from year to year calculated. This was done by looking up in Dr. Heinis' table the mental age at different chronological ages of a pupil with the given H.C., then the IQ was calculated based on the mental age thus obtained and the chronological age. The changes that would result in the IQ are shown in Fig. 4. It has not been proven that the intelligence quotients of bright children, when calculated by the standard method remain on the average constant, but there is no evidence to indicate that the very high IQ's increase to such an extent as these curves show that they must if the H.C. were constant. A child that had a constant H.C. at 110 would have an IQ of one hundred seventeen at the age of six, which would increase to one hundred thirty-four at twelve years and one hundred eighty-five at fifteen; a child with a constant H.C. of one hundred fifteen would increase in IQ from one hundred twenty-six

[1] Riley, Gordon: "A comparison of the Personal Constant and IQ." Psychological Clinic, February, 1930.

at six years of age to one hundred seventy-nine at twelve years and would be up in the four hundreds by fifteen years.

The extent by which the low IQ's must decrease if it were to be assumed that the Heinis Constant were constant is from ten to twenty

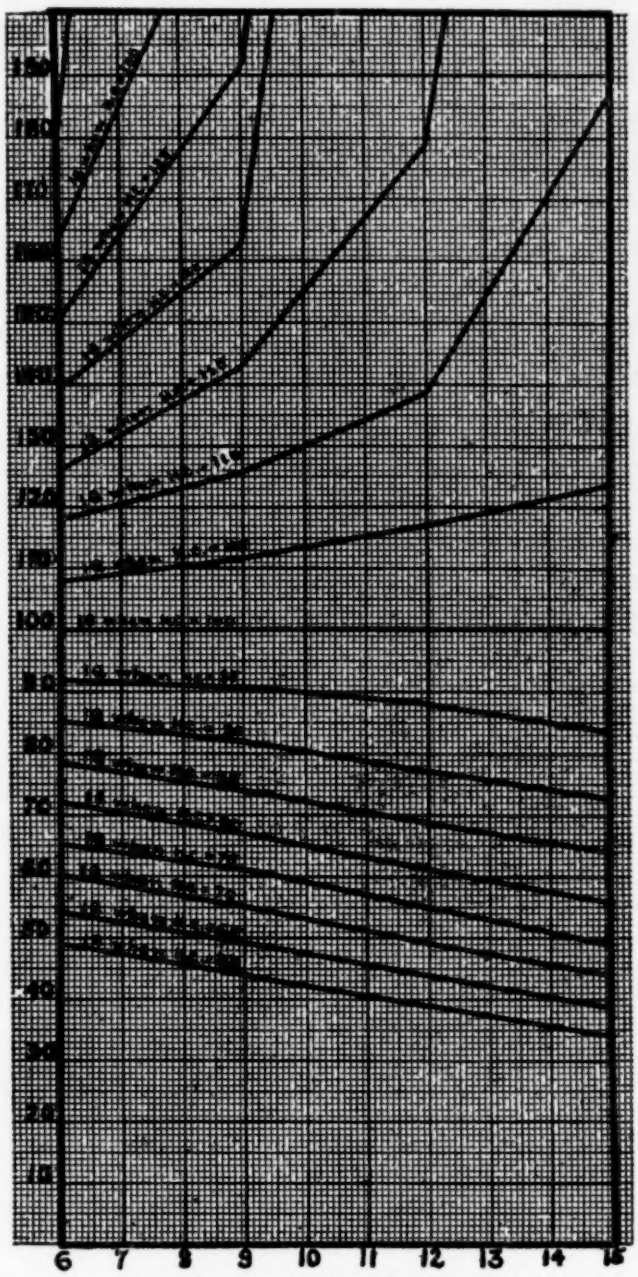


FIG. 2.—Changes that would take place in the IQ if the Heinis Constant were constant.

points as indicated in the lower part of Fig. 2. This is not beyond the realm of reason. The dull group in the present study decreased eleven points in IQ between the ages of about seven and fourteen.

Dr. Heinis put his mathematical conclusions to a test using the results of Binet tests on feeble-minded children reported by Dr. Kuhlmann.[1] He found the facts to fit with his theoretical conclusions, but unfortunately no such test was made with superior children. Our results also indicate that the Hienis Constant may be constant for children of subnormal children. The large discrepencies lie at the other end of the scale.

The present study points to the conclusion that: (1) The Heinis Personal Constant is more constant than the IQ for children below average in intelligence, at least when based on linguistic test, but that for children of above average intelligence it is less constant than the IQ. (2) There is some indication that the gain in Stanford-Binet IQ of superior pupils with age may be a function of the particular test rather than of the IQ technique.

---

[1] The Results of Repeated Mental Re-examinations of 639 Feeble-minded over a Period of Ten Years. *Journal of Applied Psychology*, Vol. V, pp. 195–224.

THE DIFFICULTY OF A MULTIPLE CHOICE TEST ITEM

PAUL HORST

Personnel Research Division, Proctor and Gamble Company

We shall define the difficulty of a multiple choice test item as being a function of that proportion of individuals answering the item which knows which of the alternatives is the best answer. This definition involves the assumption that there is some objective criterion which determines that one particular alternative is a better answer to the item than any of the others. The nature of this criterion and the logic involved in the technique for constructing it will not be discussed in the present article.

In the following analysis we simplify our problem by assuming that the text of the item involves a single idea, and that the alternatives contain the same idea to varying degrees. We assume further that the group of individuals answering the item is distributed according to some definite analytical frequency function with reference to ability to discriminate between varying degrees of the idea. That the text may contain more than one distinct idea and that each alternative may contain the same ideas in varying degrees is a condition which exists in many cases. Thus it is possible that with reference to one idea in the text the alternatives may take one order when arranged serially according to the degree to which they express that idea, while with reference to another idea they may take quite another order.

Further, it is easily possible that the group of individuals answering the test item when distributed according to ability to discriminate between varying degrees of one idea may take quite another distribution when distributed with reference to ability to discriminate between varying degrees of another idea.

Suppose now we have a multiple choice item which involves essentially but one idea. We have n alternatives which express the same idea to a greater or less degree. We represent each alternative by the amount of discriminatory ability required to differentiate between the idea expressed in it and the idea expressed in the text. These we indicate by x_1, x_2, to x_n where x_1 requires the least discriminatory ability and x_n the most.

Now suppose we have a group of individuals distributed according to ability to discriminate between varying degrees of the idea expressed in the text. Let us represent this distribution by $f(x)$.

Then none of the individuals below x_1 will be able to differentiate between the ideas expressed in any of the alternatives and the idea expressed in the text. This proportion of individuals may be represented by $\int_{x_0}^{x_1} f(x)dx$ where X_0 is the lower root of $f(x)$. Let us indicate by $S_1, S_2, \ldots S_n$ the proportion of individuals marking each alternative. We assume that all individuals below x_1 in ability will be equally likely to mark any of the alternatives, so that the proportion below X_1 marking each alternative will be $\frac{1}{n}\int_{x_0}^{x_1} f(x)dx$. The proportion of individuals between x_1 and x_2 in ability is $\int_{x_1}^{x_2} f(x)dx$ and these will be equally likely to mark any of the alternatives above x_1 so that of this proportion $\frac{1}{n-1}\int_{x_1}^{x_2} f(x)dx$ will mark each alternative from x_2 to x_n. Similarly, of the proportion between x_2 and x_3, $\frac{1}{n-2}\int_{x_2}^{x_3} f(x)dx$ will mark each alternative from x_3 to x_n. Thus we have,

$$\left.\begin{aligned}
S_1 &= \frac{1}{n}\int_{x_0}^{x_1} f(x)dx \\
S_2 &= \frac{1}{n}\int_{x_0}^{x_1} f(x)dx + \frac{1}{n-1}\int_{x_1}^{x_2} f(x)dx \\
&\cdots\cdots\cdots

x_n is the best answer. This distinction, while not of much practical importance, is of considerable theoretical importance.

If we subtract the next to last equation in (1) from the last equation, we get,

$$S_n - S_{n-1} = \int_{x_{n-1}}^{x} f(x)dx \tag{2}$$

which means that the proportion of individuals who know the best answer is the difference between the highest proportion marking an alternative and the next highest proportion marking an alternative.

Since $S_n - S_{n-1}$ gives a definite numerical value x_{n-1} can be solved for in (2) providing $f(x)$ is given. Thus the difficulty of a multiple choice test item is a function of the degree of discriminatory ability required to differentiate between the idea expressed in the text and the alternative marked by the next to highest proportion of the group.

If the alternative which the largest proportion of the group mark is not, according to certain objective criteria, the best answer, then we must assume that the text contains several ideas and that the group is not discriminating with reference to the idea for which the best answer has been objectively determined. This situation requires a more detailed analysis which will be set forth in a separate article.

Now if we have a very large number of cases, we may define our unit of measurement of discriminatory ability so that the distribution is normal, and adjust this unit so that it becomes the standard deviation of a normal distribution. If we adopt as the origin of the discriminatory scale the mean of the group, we may write (2)

$$S_n - S_{n-1} = A \int_{x_{n-1}}^{\infty} e^{-\frac{x^2}{2}} dx \tag{3}$$

and from a table of the normal probability integral we may find the value of X_{n-1} in (3). This gives us the difficulty of the test item measured from the mean of the group in terms of its standard deviation.

It should be noted from (1) and (2) that the proportion of individuals which marks the best answer by chance is the same as the proportion which marks the next best answer. From this it is clear that for a test item of any given degree of difficulty the proportion marking the best answer will contain the least possible proportion due to chance when all alternatives but the best answer are of equal difficulty, for with the proportion marking the best answer remaining constant, the

proportion marking the next best answer must be reduced as the other alternatives approach it in difficulty.

This fact is of considerable importance in constructing multiple choice test items, for it is evident that fewer alternatives, equal in difficulty, may result in a smaller chance element in the answer than is given by a greater number of alternatives presenting a wider range of difficulty.

COMMUNICATIONS AND DISCUSSIONS

IN DISAGREEMENT WITH E. A. LINCOLN'S ARTICLE "THE UNRELIABILITY OF RELIABILITY COEFFICIENTS"

LUTON ACKERSON

Institute for Juvenile Research, Chicago

Professor Lincoln's article with the above provocative title[1] has already been discussed by Franzen and Derryberry.[2] While agreeing entirely with these critics, I feel that more should be added, with the aid of further figures from Lincoln's data which he has generously sent me in full knowledge that I would endeavor to use them to refute his thesis. He writes that "the median difference for boys' sternal height is 6.2 mm., and the reliability coefficient is .994, while the median difference for boys' head width is 1.7 mm., with a reliability coefficient of only .591. The *same inconsistency* [italics mine] is found in the results for the girls. Here the median difference in sternal height is 5.0 mm., with a reliability of .995, while in head width the median difference is 1.7 mm., and the coefficient is .694" (p. 13). He concludes that since the higher reliability coefficient is accompanied by larger *absolute* differences between two series of measures in his material there must be something wrong with reliability coefficients as a device for demonstrating the agreement between parallel series of observations.

His error lies in attempting to compare absolute *raw* values in different fields of data without regard to the respective variabilities of these fields, as Franzen and Derryberry point out. The most satisfactory method of rendering different series of measures comparable is by reducing the original scores to "standard measures," a procedure long known to statisticians and adequately described in Kelley's[3] and Holzinger's[4] texts, but unfortunately still a custom of the future so far as most of us biometricians and psychometricians are concerned. Although Lincoln's data were not presented in the form of standard measures, we are safe in inferring that by dividing his median differ-

---

[1] *Journal of Educational Psychology*, Vol. XXIII, 1932, pp. 11–14.

[2] R. H. Franzen and C. M. Derryberry: *Journal of Educational Psychology*, Vol. XXIII, 1932, pp. 559–560.

[3] T. L. Kelley: "Statistical Method." New York: Macmillan, 1923, pp. 114–117.

[4] K. J. Holzinger: "Statistical Methods for Students in Education." New York: Ginn, 1928, pp. 118–122.

ences by the appropriate standard deviations, we may obtain a fair estimate of what the results would have been if standard measures had been used.

Among his seventh-grade boys the standard deviations of sternal height and head width were about 72.0 mm., and 5.7 mm., respectively.[1] In terms of standard measures, then, the median or average differences between the two sternal-height series and between the two head-width series are not far from .09 and .30, respectively. That is, the discrepancy between the two series of head-width measures is about three times as great as that between the two series of sternal-height measures. When we recall that the reliability coefficient for head width was only .591 while that for sternal height reached the high value of .994, we see that, after all, the reliability coefficient gives a truer picture of agreement or disagreement then does the mere comparison of absolute "actual differences," as Lincoln terms them.

Among his girls the reliability coefficient similarly provides a more meaningful index of the adequacy of the measurement-taking. The standard deviations for sternal height and for head width were about 72.0 mm. and 8.5, respectively, and the median or average differences in terms of standard measures would be approximately .07 and .20, respectively, while the corresponding reliability correlations were .995 and .694. Here again, the reliability coefficients serve as the better indicator of the accuracy or consistency with which these traits were measured. In other words, head width in his data was being measured much less adequately than the vertical anthropometric measures, in spite of the smaller "actual differences" which were found in the head-width series.

In fact, to follow Lincoln's criterion that the degree of reliability of two series of measures may be inferred from their actual differences, which in a statistical sense are only arbitrary numerical values, would lead to absurd conclusions. For example, nasal length measured in centimeters would thereby be a much more reliable measure than total stature measured in centimeters, regardless of how clumsily nasal length may be measured, merely because the figures representing the raw differences would be smaller. Or similarly, one could argue with equal correctness or incorrectness that measurements of stature when recorded in inches would be more reliable than when recorded in centimeters because the actual raw differences were smaller.

---

[1] From a personal communication from Professor Lincoln.

In short, the present writer's perusal of Professor Lincoln's data has not altered his belief that there is no substitute for reliability correlations in studies based upon variate data.

RELIABILITY COEFFICIENTS ARE STILL UNRELIABLE

EDWARD A. LINCOLN

Harvard University

After the publication of an article entitled *The Unreliability of Reliability Coefficients*,[1] the writer received several letters pointing out alleged fallacies in his argument. Also, two or three articles have been published which take the same stand.[2] Perhaps another brief word will be of interest to the critics and others who are concerned with this problem.

The chief contention of the critics is that the reliability coefficient is a better measure because it is a *standard* measure; that is, it is independent of the absolute raw values of the measures taken. The latter part of this statement is, of course, true. It is also true, however, that in some instances the absolute rather than the relative values are important to the investigator, and then a relative measure is neither valid nor useful. That was the case in the illustration cited in the original article, and the same argument would apply in many other cases.

It should be remembered that the same characteristic of the reliability coefficient which makes it very serviceable in some instances renders it misleading and worthless in others.[3] This, as every real statistician knows, holds true for every statistical measure which we have. That is, no measure is always and inherently better than any other; it is only better to show some things.

Let us consider again the problem with which the original article dealt. The writer was seeking an answer to the question, Is the technique of measuring heights employed in the Harvard Growth

---

[1] *Journal of Educational Psychology*, Vol. XXIII, 1932, p. 11–14.

[2] Franzen, R. H. and C. M. Dewberry: *Journal of Educational Psychology*, Vol. XXIII, 1932, pp. 559–560.

Heilman, J. D.: *The Teachers Journal and Abstract*, Vol. VII, 1932, p. 93.

Akerson, L.: *Journal of Educational Psychology*, Vol. XXIV, No. 3.

[3] Heilman makes this point very aptly. He says, "No one misinterprets the coefficient of reliability who remembers that it is intended to measure only variable errors and that it varies with the standard deviation of the distribution."

Study giving measurements of the individual pupil which are accurate and dependable? The reliability coefficients of .99 plus said, Yes. When, however, a study was made of the actual differences in two sets of measurements it was discovered that in nearly half the cases there was a difference of half a centimeter, and that there were many differences of a whole centimeter or more. Clearly the reliability coefficients in these cases were very misleading, and no one has successfully refuted this important point. In short, the critics have explained why the difficulty occurs, but they have not explained it away. The writer has neither seen nor heard any argument which moves him to change his original position that reliability coefficients when used in such a study are very unreliable.

STUDENTS' RATINGS OF LECTURES
1. COMPARATIVE RATING OF GENERAL PSYCHOLOGY LECTURE PARTS

NATHAN ISRAELI

University of Maine

AN INTROSPECTIVE STUDY

Not all parts of a lecture appeal alike or are equally interesting. Just how students rate comparatively lecture parts on the basis of interest can be measured. This study is based on subjective or introspective judgments of students. Incidentally, it leads to a comparison of interest ratings of textbook and non-textbook lecture material. Demonstrations, delineoscope illustrations, material on the textbook, non-textbook material, data from the lecturer's experiments are five parts of lectures in a General Psychology course rated by students for interest.

A LECTURE RATING QUIZ

Two different classes in General Psychology were given the following lecture elements rating quiz:

This quiz is for pedagogic and scientific purposes. Make your answers express your true opinion.

Here are five parts of the lectures in this course. You are to number each item in accordance with your opinion of how interesting

they are. The most interesting item should be numbered 1, and the least interesting item should be numbered 5.

Write Number Here	Items of the General Psychology Lectures
———	*A*. Demonstrations[1]
———	*B*. Delineoscope illustrations
———	*C*. Lecture material on the textbook
———	*D*. Non-textbook lecture material[2]
———	*E*. Lecture material from the present lecturer's own experiments, etc.

Group *A*, sixty-six subjects, and Group *B*, thirty-eight subjects, included mostly sophomores, had the same instructor, and were not equivalent groups.

RESULTS

The most interesting part of a General Psychology lecture consists of demonstrations. Both groups think this is so. For Group *A*, next in order of interest are the delineoscope illustrations, non-textbook lecture material, textbook material, and lecturer's illustrations; and for Group *B*, lecturer's illustrations, non-textbook, textbook material, delineoscope. This is shown in Table I.

Table I.—Most Interesting Lecture Element

Lecture-part	Percentage rating a lecture-part as the most interesting one[1]	
	Group *A*	Group *B*
A. Demonstrations	65.2	57.9
B. Delineoscope illustrations	10.6	2.6
D. Non-textbook lecture material	10.6	5.3
C. Lecture material on textbook	9.1	10.5
E. Material from lecturer's experiments	4.5	23.7

[1] Taken from the first of a five column order-of-merit table for each group. Data in this table are taken from the first column of order-of-merit tables not included here.

Another way to look at the results is to take the ranking item in each of the five columns of the order-of-merit table and we find the

[1] For example, the demonstration of performance tests, demonstration of individual differences in emotional reactions to music, etc.

[2] Part of the lecture covering material not in ——— or not discussed by ——— in detail.

following order for Group *A*: demonstrations, delineoscope, lecturer's material, lecturer's material, non-textbook. For Group *B*: demonstrations, delineoscope, delineoscope and textbook, non-textbook, textbook. According to this series of ranking items in each of a five column order-of-merit table, demonstrations rank as the most interesting lecture element.

Weighting each column in the order-of-merit table gives a combined score from the different columns for each one of the five lecture items. Column one is multiplied by five, two by four, three by three, four by two, and five by one. Total scores are computed for the different lecture items on the basis of frequency and not percentage. The total scores, given in the first two columns of table IV, are reduced in the last two columns to a comparable basis, based on lecture

TABLE II.—COMPARATIVE RATING OF GENERAL PSYCHOLOGY LECTURE PARTS

Lecture-part	Total score		Rank order		Comparative rating	
	Group *A*	Group *B*	Group *A*	Group *B*	Group *A*	Group *B*
A. Demonstrations	283	151	1	1	100	100
B. Delineoscope illustrations	185	100	2	4	65.4	66.2
E. Lecturer's illustrations	183	127	3	2	64.7	84.1
D. Non-textbook lecture material	172	112	4	3	60.8	74.2
C. Lecture material on textbook	167	80	5	5	59.0	53.0

elements *A* (demonstrations) as one hundred. The middle columns show the new rank order of the General Psychology lecture parts for each group.

In all instances, practically, the non-textbook ranks higher than the textbook material. To contrast further these two important lecture parts, the total scores of part *E*, lecturer's material, and part *D*, non-textbook material, are combined together and the new total score may be compared to that of part *C*, the textbook material. The new total score for the combined non-textbook lecture material is three hundred fifty-five for Group *A* and two hundred thirty-nine for Group *B*. Or, the combined non-lecture material is 2.13 times as interesting as the textbook material to Group *A* and is 2.99 times as interesting as the textbook material to Group *B*.[1]

[1] This treatment of the data suggests further work on a comparison of textbook and non-textbook material. In a different experiment, it was found that Organic

SUMMARY

Five parts of a General Psychology lecture courses were rated for their interest by two different sections in that course. The most interesting part of the lectures consists of demonstrations and the least interesting part, of the textbook material. Combined non-textbook material is 2.13 and 2.99 times as interesting as the textbook material to the two different classes.

Chemistry students rate the lecture on the textbook material as less valuable for instruction than the lecture on non-textbook material.

BOOK REVIEW

The Vibrato. Iowa City: University of Iowa Studies in the Psychology of Music, New Series No. 225, April, 1932.

This is a thorough and excellent piece of research work in a field much in need of experimental investigation. Following closely the pioneer experimental and more localized work of Cheslock,[1] it unravels further the so-called mysteries of this "soul of the musician's tone." The University of Iowa volume, edited by Carl Seashore, is a collection of studies by eleven investigators. By far the greater part is devoted to voice; in fact, the entire volume could perhaps on this account have been more appropriately titled Voice Vibrato.

The effects of pitch, intensity and periodicity, the three components of vibrato and their interesting and significant inter-relations are analyzed in detail with ample diagrammatic illustrations. The findings show that fluctuations in these components are intimately linked with our concept of timbre and quality; that the vibrato can be modified through training by directing the attention to certain aspects of the problem; that vibrato itself varies considerably with the tonal and emotional environment in which it is used. The very careful methods of procedure used and the specific nature of the individual problems give the results dependability and conciseness. In addition there is a historical summary of established facts in experimental studies on vibrato brought up to date, detailed description of phonophotographic apparatus used, comparative study of student vibrato and artist vibrato, of concert singing and opera singing, of attack, release and transition in vibrato. The volume contains a wealth of significant and interesting material. A condensed, non-technical summary of these studies will be helpful to music students and teachers not versed in psychological technique and procedure.

OTTO ORTMANN.

Peabody Conservatory of Music, Baltimore.

[1] Cheslock, Louis: "Introductory Study on Violin Vibrato." Peabody Conservatory Research Studies.